

Technology transfer requires an entrepreneurial academia

Attempts to replicate the recent success of US technology transfer need to focus on the toughest elements of the problem: the availability of venture capital and the structure of the university system.

Finding the magic formula for the successful transfer of technology from universities and government laboratories to commercial application is a perennial quest of governments around the world. With the United States enjoying an economic boom of unprecedented strength and duration — partly on the back of its success in information technology and, to a lesser extent, biotechnology — interest in the question is all the greater.

This summer, for example, Japan's parliament has passed a law, similar to the Bayh–Dole Act passed in the United States in 1980, that will allow universities and government laboratories to grant exclusive patents to private companies with which they have gone into partnership (see page 3). An influential panel in Canada, meanwhile, is urging the government to strengthen universities' means of protecting and exploiting their intellectual property, and to push the universities to pursue commercial opportunities more aggressively (see *Nature* **400**, 805; 1999).

Various factors have contributed to the US success, and some are easier to replicate than others. A legal framework has been put in place which allows private companies in the United States to gain exclusive patent rights to ideas that were initially developed with public money. An argument can be made that this is unfair to the taxpayer. But experience suggests that only the private sector is ultimately capable of efficient commercial exploitation of such ideas. It follows that, elsewhere as in the United States, public research agencies must accept the need to pass the baton of intellectual-property rights

firmly and unconditionally to the private sector.

The availability of venture capital is also critically important, but it is not something that governments can readily legislate for. Venture-capital markets in Europe and elsewhere remain disappointingly small compared with that of the United States, whose diversity of venture-capital sources — including wealthy individuals willing to take a gamble on a good idea — remains unmatched elsewhere.

Perhaps the most impressive element of the US technology-transfer chain, however, is the university system itself. Public and private universities are able to compete freely in an environment in which faculty are highly mobile and strong departments can rapidly emerge — or decline — subject to the quality of their work. The system is always ready to take a chance on talented young faculty, and is itself highly entrepreneurial.

That environment is hard to replicate alongside the state-controlled higher-education system of other industrialized countries. Most of them have tried to shake things up over the past decade, with conspicuously little success. Powerful interests, most notably senior academics, are deeply entrenched and opposed to change.

A legal structure that allows intellectual property to escape from universities and government laboratories is a necessary but insufficient prerequisite for efficient technology transfer. It is the absence of venture capital and, above all, of a dynamic and flexible university system that is most likely to frustrate those seeking to emulate US strength in technology transfer. ■

Changing *Nature* (cont.)

This issue includes some enhancements to *Nature's* subliminal impact.

Editors of any publication are well used to the fact that most readers simply want to get the 'story' — the news, the opinion, the new discoveries — whereas the skills that are deployed in writing, editing and visually presenting the story are, by and large, taken for granted. That, of course, is a sign of professional success — readers are quick to notice when standards slip, but may not fully appreciate the reasons why an excellent piece of content achieves its impact. Never mind. Informal feedback and the harder measure of climbing circulation suggest that, although they may not appreciate the full reasons for it, the number of people wanting or needing to open *Nature's* pages every week continues to grow.

Those discreet aspects of content are more noticeable when they are changed, however, and several such developments are marked by this week's issue. Take our design. In January last year we introduced a significant change in our appearance. That redesign has withstood the test of time. Nevertheless, we feel that some slight improvements are desirable, which is why the pages in the front sections in particular now have a lighter feel to them, stemming from changes in typeface, the use of linear design features and the style of page layout.

We have also changed the order of sections. Time was when book

reviews were published at the back of the journal. More recently they have been sandwiched between two much more technical sections. The Book Reviews section surely deserves more prominence. It now appears further towards the front.

And finally, we have changed the name of the Scientific Correspondence section to 'Brief Communications'. The primary reason is to avoid confusion with our Correspondence section, which consists of informal communications on non-technical matters. As with Scientific Correspondence, Brief Communications are peer reviewed and scientific in character. Some are responses to papers published in *Nature*. The majority, however, report self-contained scientific results that are of unusual interest — perhaps for their policy implications, or for their unexpectedness, or even for their sheer entertainment value. Although the advance in fundamental understanding represented by Brief Communications may not typically be as great as in our Articles and Letters, the standards of peer review that we apply to their technical quality are no less strict, while the interest aroused will typically span a broad range across the many disciplines of our readership.

Further changes are in the pipeline. Meanwhile, we hope readers' enjoyment of *Nature* will be at least subliminally enhanced. ■



**AIDS**

Governments urged to back research into an AIDS vaccine

p4

**San Andreas**

Observatory drillers seek insights in the San Andreas fault

p5

**Synchrotron**

Middle East tour aims to raise support for threatened project

p6

**Instruments**

Ageing equipment is slowing down US biologists

p7

Japan acts to speed technology transfer from universities

Tokyo

A Japanese law that takes effect this autumn will promote technology transfer from universities and government laboratories to the commercial sector.

The technology-transfer bill, based on the influential Bayh–Dole legislation passed almost 20 years ago in the United States, is part of a policy initiative aimed at improving links between government, industry and academia. It underpins prime minister Keizo Obuchi's Millennium Project, which aims to foster new industries in areas such as biotechnology, information technology and environmental engineering.



Obuchi: fostering new industries.

Obuchi's project aims to make industry competitive by increasing collaboration with government-run research institutes. It will step up support for genome-related research, including a plan

to decode a third of human gene sequences and to create a national centre for bioinformatics by 2001 (see *Nature* 400, 389; 1999).

Industry has welcomed the new framework. Until now, rights to patents produced from government-funded research have belonged to the national treasury, and could not be assigned to the private sector without complex legal procedures. The new measures will encourage universities to collaborate with industry to promote the application of government-funded science — allowing technology licensing organizations (TLOs), which advise on the commercial exploitation of university research, to obtain patent rights to inventions made at universities.

The technology-transfer law will grant awardee organizations first rights to a patent fully or partially funded by the government, allowing industrial research partners to retain title to inventions.

The Ministry of International Trade and Industry (MITI) says the licensing income of US universities alone totalled US\$0.7 billion in 1998, while the combined licensing income of Japan's universities and national

research institutes is only US\$2.7 million.

"University research is potentially a large source of income, as well as new technology," says Masahiro Hashimoto, head of industry–university collaboration at MITI. "But the main problem is the reluctance of university researchers to become involved in profit-making activities."

Japan's civil service law forbids commercial activity by employees of national universities, and bans academics from holding external posts at private companies (see *Nature* 399, 624; 1999), though the government is planning to relax these restrictions.

"Japanese universities seem to be allergic to commercial activities: many still see involvement in such activities as a sign of frivolity," says Takeshi Onoda of the Japan Federation of Economic Organizations, the former director of Mitsubishi Chemicals.

The transfer of university research will bring immediate benefits to the commercial sector, but not necessarily to the universities themselves, says Onoda. "Successful application of the new technology-transfer law depends on whether individual researchers will be able to promote the commercialization of their inventions, and how mediating organizations such as TLOs succeed in assisting such endeavours."

Although Obuchi's Millennium Project envisages the emergence of new industries over the next ten years, some observers predict that university technology transfer in Japan will take up to 15 years to develop.

"There is a need to develop a proper infrastructure for collaboration between industry, academia and the government, and we are 20 years behind the United States in this area," says Onoda.

Asako Saegusa

Crop trials seek to allay public fears

Tokyo

Japan's Ministry of Agriculture, Forestry and Fisheries (MAFF) has announced plans for a five-year project to examine potential long-term risks of genetically modified (GM) crops to the environment. The move comes amid growing signs of consumer resistance to GM products in Japan.

Researchers at the National Institute of Agro-Environmental Sciences, in Tsukuba, will test the possibility of gene transfer from GM oilseed rape to wild plant species.

GM oilseed rape, approved in Japan for agricultural purposes, is an outbreeding species with many wild relatives in Japan. If gene transfer is judged likely, MAFF will assess whether this would have any harmful ecological effects.

According to MAFF, results from this research will be reflected in its new safety guidelines, which are currently being revised to incorporate more rigorous criteria for testing the safety of GM crops. This is partly due to the recent publication of research indicating that pollen from *Bt* corn can harm the larvae of monarch

butterflies. The ministry decided in June to suspend approval of all *Bt* crops for agricultural purposes until it has established criteria for evaluating the safety of GM crops (see *Nature* 399, 719; 1999).

Although MAFF officials predict that the risks of GM crops are negligible, they emphasize the importance of allaying public concern by tightening the safety regulations.

In the wake of last month's decision by the government to label food products containing detectable GM ingredients (see *Nature* 400, 605; 1999), Japanese companies — none of which has yet commercialized GM products — are finding it hard to overcome the negative image of GM foods.

Recent government decisions on GM organisms "are a real setback for Japan's GM industry," says Atsushi Suzuki, an adviser to Japan Tobacco, which is involved in the development of GM rice.

Some companies are already retreating. Kirin Beer, one of Japan's largest breweries, last week announced it would abandon its research into GM tomatoes, and promised its beer would be GM-free by 2001.

A. S.

Cash bonuses won't stop brain drain, say Canadian academics

Montreal

Quebec's provincial government may offer cash bonuses to researchers and academics, to try to stop them being wooed away by US institutions.

The provincial cabinet has formed a panel to study the brain drain. But research minister Jean Rochon, who leads the study group, says cash bonuses are not the only answer. He believes "a basket" of conditions is needed to keep researchers in Quebec.

Arpi Hamalian, president of the Quebec University Professors Federation, says bonuses won't compensate for years of university underfunding. A strategy to stop the brain drain should "restore the collegial environment that allows for good teaching and research; that means restoring funding," says the Concordia University education professor.

A study by the independent Conference Board of Canada says that official figures including only permanent emigrants underestimate the problem.

"Until 1989, the combined number of permanent and non-permanent Canadian emigrants to the United States was fairly stable at 17,000 annually," says the board. "By 1997, the number had increased to 98,000, including a six-fold increase by non-permanent emigrants."

Researcher Mahmood Iqbal says the North American Free Trade Agreement changed the picture so that more than 90 per cent of US visas granted to Canadian professionals are now non-permanent, and that up to 60 per cent of the people holding them elect to stay permanently.

Prime Minister Jean Chrétien argues that the emigration of professionals is matched or exceeded by the inflow. But Iqbal says these are not comparable in quality, and it may take up to ten years for incomers to equal their Canadian counterparts' productivity. **David Spurgeon**



Lucky shot? A gp120 trial in Thailand is currently the world's only phase 3 AIDS vaccine trial.

Call for governments to back development of AIDS vaccine

Erice, Sicily

Scientists around the world will be asked to endorse a statement underlining the urgency of developing an AIDS vaccine and calling on governments — especially of poor countries — to be more actively involved.

The statement was drawn up at a two-day World Federation of Scientists workshop at the Ettore Majorana Centre for Scientific Culture in Erice, Sicily. Signatories already include two of the world's best-known AIDS researchers — Robert Gallo of the Institute of Human Virology in Baltimore, and Luc Montagnier, director of the World Foundation for AIDS Research and Prevention in Paris.

The workshop was told that vaccine development has consumed only 1 per cent of the US\$18 billion that has been spent fighting AIDS, and that only one full-scale phase 3 clinical trial is under way, with a vaccine using the gp120 protein in Thailand. A second trial is about to start in Uganda.

"Research into vaccine development has been disproportionately low compared to the potential public health benefits," says Seth Berkley, head of the New York-based International Aids Vaccine Initiative, which has been working to attract funds from bodies such as the World Bank and the Bill and Melinda Gates Foundation. "Even though research is now well supported, product development is not," he adds.

Scientists from poorer countries agree. "We need to get politicians on board at the highest level," says Fred Mhalu, head of Tanzania's anti-AIDS programme.

"In Brazil, AIDS activists have convinced the government to spend \$600 million on AIDS therapies," says Mauro Schechter, of the University of Rio de Janeiro. "But we still have difficulty in convincing officials of the

need to provide funds for research on vaccines. There is still this feeling that vaccines are developed in the industrialized countries, whereas there is much that countries like Brazil can do, for example preparing the ground for efficacy trials, and perhaps eventually producing the vaccines themselves."

The authors of the Erice statement say it will help to increase awareness not only of the urgency and scientific feasibility of AIDS vaccine research, but of the many steps that developing countries can take.

Reflecting concern that some scientists may be reluctant to move into clinical trials without being confident of high levels of success, the statement says that the urgency "is such that even a partially effective vaccine would still be valuable in significantly reducing the rate at which HIV is spreading".

Recognizing that the biggest demand for a vaccine will be in developing countries — so that it would be unlikely to produce substantial profits for manufacturers — the statement calls for new incentives to draw companies into public/private partnerships.

The authors of the statement hope that, if widely endorsed, it will help to increase political commitment to the national AIDS vaccine programmes that are now starting up in countries such as China, South Africa, Russia and Brazil, where the epidemic continues to grow rapidly.

They also hope to stimulate support from industrialized nations for international efforts, such as that led by the World Bank — whose board will discuss a proposal for increased support of AIDS vaccine development later this month — and a joint initiative between UNAIDS and the World Health Organization, which is due to be launched early next year.

David Dickson

Go-ahead for San Andreas drilling project ...

San Diego

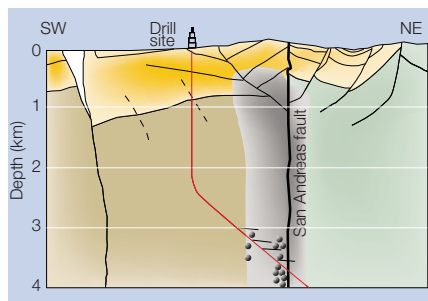
An international group of scientists is set to win support from the US National Science Foundation (NSF) for a long-planned project to drill deep into California's San Andreas fault in search of better data on seismic activity.

The San Andreas Fault Observatory at Depth is designed to give scientists a new insight into the physical and chemical processes controlling the mechanics of the fault.

The observatory project was approved in late July by the influential Programs and Plans Committee of the National Science Board — the NSF's governing body — which approved \$16 million for equipment. The full board will consider the recommendation in November.

If, as is likely, the board accepts the recommendation, the project will move through the congressional funding process, where an estimated additional \$10 million will be sought for six years of associated research. Drilling of the observatory well — at an isolated site in the coastal mountains about halfway between San Francisco and Los Angeles — is due to begin in late 2001.

Stephen Hickman, a co-principal investigator from the US Geological Survey's earthquake hazard team in Menlo Park, California, says: "One of the more exciting aspects will be the ability to look at the nucleation process of



Boring science: geologists hope to probe the San Andreas fault with a well four kilometres deep.

earthquakes, particularly the aseismic phases before an earthquake takes off."

The San Andreas observatory is part of a series of NSF initiatives called EarthScope. This includes a project called USArray, which deploys mobile seismometers to map the Earth's structure under the United States.

The Programs and Plans Committee approved \$58 million for equipment for USArray in July, and the full science board will decide on this project in November. A plate-boundary observatory is also under discussion, but this is not slated for funding consideration until much later.

"All of these projects are designed to give us a better understanding of the physics of earthquakes," says James Whitcomb, acting deputy director of the NSF's Earth science division.

The San Andreas observatory involves 33 principal investigators from 19 US universities, 15 scientists from the US Geological Survey, and research teams from 12 institutions in Japan, France, Germany and Britain.

Planning for the observatory began in 1992, although the concept of drilling into a fault has been debated for decades. Initially, the scientists planned to drill a ten-kilometre-deep observatory, but cost and technical difficulties meant they decided on a four-kilometre-deep well.

Seven locations along the San Andreas fault, between the San Francisco area and east of Los Angeles, were considered for drilling. The team chose Parkfield, where the US Geological Survey has long had a ground monitoring seismology station.

The team's proposal says that an important area of investigation is the "evidence indicating that the slip in crustal earthquakes along major plate-bounding faults — like the San Andreas — occurs at extremely low levels of shear stress." It adds that knowing "how crustal faults lose their strength is critically important" in the mechanics of earthquakes and hazard reduction.

"Our current knowledge of the fault zone process is so poor that not only are we unable to make reliable, short-term earthquake predictions," the proposal says, "we cannot scientifically assess whether or not such predictions are even possible."

Rex Dalton

... and more money for the Earth sciences in Germany

Munich

The German science ministry has launched an Earth sciences programme that will give priority to interdisciplinary projects with socio-economic relevance.

Germany's research minister, Edelgard Bulmahn, has allocated DM500 million (US\$270 million) over the next 15 years for funding the new programme, called Geotechnologies, which will start next year.

In addition, the Deutsche Forschungsgemeinschaft, Germany's main funding agency for university research, will make about DM20 million per year available to support interdepartmental university research in the Earth sciences.

The money is urgently needed: geological research in Germany has been in crisis since the mid-1990s, when several large-scale projects, such as the continental drilling programme, came to an end (see *Nature* 382, 196; 1996).

The new programme is designed to unite the fragmented German Earth science community and increase the relevance of

their work to the management of resources and natural disasters.

Among the themes chosen for funding are global climate change, the tectonics of continental margins, prospecting of gas hydrates, variations in geomagnetism, and early warning systems for earthquakes, floods and volcanic eruptions.

The programme will promote collaboration between universities and other institutes, such as Germany's National Research Centres, which can also apply for funds.

Rolf Emmermann, the scientific director of the Geoforschungszentrum (GFZ) in Potsdam, Germany's largest geological sciences institute, is optimistic that what he calls the "exemplary integrative approach" of the new scheme will keep German researchers at the forefront of international Earth sciences.

The social relevance of today's Earth sciences is not in doubt, says Emmermann. "The results of our research leave their mark in many topical political discussions about

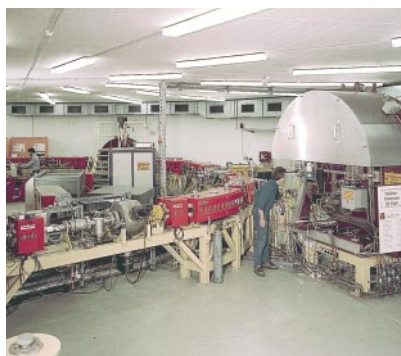
climate policy or disaster management."

In earthquake research, the GFZ is involved in developing early warning systems for some European cities in endangered regions, including Istanbul and Bucharest. Moreover, it participates in international research projects aimed at understanding the mechanisms of the tectonic processes — including the assumed role of fluids in the Earth's crystalline crust — that generate earthquakes.

Emmermann hopes that the interdisciplinary approach of the new programme will encourage collaboration between seismologists and engineers in setting building codes for earthquake regions.

With this in mind, the GFZ has already begun analysing the damage caused by earthquakes in Turkey and the Middle East. Earthquake experts have blamed the disastrous consequences of the recent quake in Turkey on the lack, or disregard, of such standards (see *Nature* 400, 803; 1999).

Quirin Schiermeier



Sesame seed: Berlin's Bessy-1 could create a new facility in the Middle East.

Funding problems threaten Middle East's synchrotron

Paris

Scientists will tour the Middle East this month to drum up support for a planned synchrotron facility, which is threatened by lack of funding.

The Sesame (Synchrotron radiation for Experimental Science and Applications in the Middle East) project would involve moving and upgrading Bessy-1, a decommissioned 0.8 GeV synchrotron radiation source in Berlin. But it must be dismantled and packed by the end of this year. That alone is expected to cost half a million dollars.

Maurizio Iaccarino, United Nations Educational, Scientific and Cultural Organization (Unesco) assistant director-general for natural sciences, and Herwig Schopper, former director-general of the European Laboratory for Particle Physics (CERN) and president of the project's interim council, are to visit several countries that have shown an interest in the project.

Because the facility would be open to scientists of any nationality, organizers see its potential as not only a world-class research centre, but also as a politically important example of scientific cooperation in the region.

"A project costing US\$50 million over ten years needs a strong political commitment," says Siegbert Raither, director of mathematics, physical and chemical sciences at Unesco.

Egypt, Iran, the Palestinian Authority and Armenia have all made official bids to host the centre, and several other countries in the region are supportive. But they are not expected to be able to finance the venture themselves, so organizers need to seek financial aid from the European Union countries and the United States.

Heather McCabe

E-Biomed to be launched as a repository for research

Washington

A centralized database of scientific literature for biology researchers is to be launched in January by the National Institutes of Health (NIH). It will contain the content of many established journals as well as material that has not yet been peer-reviewed.

The repository is the latest manifestation of the E-Biomed electronic publishing proposal, introduced by NIH director Harold Varmus in April (See *Nature* 399, 8; 1999). Known as PubMed Central, it will be integrated with PubMed, the National Library of Medicine's existing literature database.

It will include reports that have been screened by what the NIH calls "responsible groups" but not formally peer-reviewed. "This material will be clearly distinguishable from the peer-reviewed content of PubMed Central," the proposal says. It is intended as the initial site in a system to be overseen by an international advisory committee.

After an outcry from scientific publishers, the proposal makes clear that the NIH itself will not arrange either for peer review of papers in PubMed Central, or for screening of the non-peer-reviewed section.

Most of the peer-reviewed section of the repository is expected to come from existing scientific journals, whose publishers may

make their content available after a short lag time: one scientific society says it may delay the appearance of papers from its journals on PubMed Central by about six months.

Scientists will be able to submit material directly to PubMed Central, if they form themselves into what the proposal calls "any organisation with at least three members who are principal investigators on research grants from major funding agencies".

The new proposal was released on 30 August, when neither Varmus nor David Lipman, head of NIH's National Centre for Biotechnology Information and a key author of the proposal, were available to discuss it. Several of the most strident critics of past E-Biomed proposals were on vacation.

However, Dale Benos, chair of physiology and biophysics at the University of Alabama and head of the publications committee of the American Physiology Society, predicts that the proposal will be welcomed by societies that were initially wary.

"Varmus should be congratulated — he has listened to criticism and reformulated it in a sensible way," says Benos. "There is not going to be an E-Biomed board of editors, it will let individual journals maintain their autonomy, and things will work much as they do now."

Colin Macilwain

Pollard resigns as president of Salk

San Diego

The Salk Institute has announced that Thomas Pollard will step down as president at the start of next year, ending months of speculation about leadership at the biomedical research centre at La Jolla, California.

Pollard, who had two years remaining on his five-year administrative contract, will continue as a professor at the Salk, running the institute's structural biology laboratory.

In February, the Salk's board stripped Pollard of his title as chief executive after internal debate over his leadership (see *Nature* 400, 805; 1999). The board then undertook a governance study — completed in July — and is considering how to apply its recommendations on faculty, board and administrative issues.

"I'm proud of what we've accomplished," said Pollard in a statement. "However, I have decided to step down as president so I can devote more time to my scientific research. This change also will allow me to spend more time on public service, particularly in advocacy for funding for biomedical research."

Frederick Rentschler, chairman of the Salk's board and chief executive since February, issued a statement, thanking Pollard for his "vigorous efforts" as president. Rentschler, a retired corporate executive, became chief executive after some senior faculty members became displeased with Pollard.

Other senior staff, including cell biologist Ian Trowbridge, are concerned that Pollard's ousting will make it difficult to attract another top-flight scientific leader. Asked about plans to replace Pollard, Trowbridge said: "There has been no internal discussion with the faculty on leadership." Stephen Heine-mann, chairman of the institute's faculty, declined to comment.

The Salk Institute board has created a six-person search committee to find a replacement, chaired by investment banker Jerome Kohlberg. It remains unclear how the Salk Institute's executive and scientific leadership will be organized in the future. Institute officials have not released the consultant's report, but some staff may now push to be told its recommendations.

Rex Dalton

US faces increasing demand for funds to replace ageing equipment

Despite moves to boost funding, US biologists are struggling to pay for the instruments they need to remain competitive.

Last autumn, Gerald Fischbach, then the new director of the US National Institute of Neurological Disorders and Stroke, sent an unusual letter to the institute's 2,000 extramural scientists, inviting them to apply for one-off \$50,000 grants for research equipment. The response, he says, was "amazing".

The institute was swamped with 1,800 applications, requesting tools including fluorescence microscopes, sophisticated centrifuges and chromatography equipment. "If we funded every one of them it would have cost \$60 million," says Fischbach.

Fischbach had just \$10 million available, but has since added \$5 million to the programme, allowing him to fund one in four applications. "These were not unreasonable requests," he adds. "These are people who have used the same microscope for ten years."

Tremendous problem

The institute's experience is just one measure of a burgeoning need for up-to-date equipment being felt throughout the biological science community, say university administrators, government officials and scientists.

"Just about every scientist I've ever talked to who is running a big lab is concerned about instrumentation," says Bill Brinkley, outgoing president of the Federation of American Societies for Experimental Biology (FASEB).

"It's a tremendous problem. You can't underestimate it," adds Stanley Opella, a chemist at the University of Pennsylvania, who relies on NMR machines to solve the structure of membrane proteins. There are virtually no sources of federal money for costly instruments, he says.

Warp-speed technological advances have pushed up the price of instruments, while making it vital for laboratories to keep their machinery updated to remain competitive.

Yet this is almost impossible using funds from individual investigator grants administered by the National Institutes of Health (NIH) and the National Science Foundation (NSF). "The NIH is not going to give you \$200,000 up front in an R01 grant" to help purchase a microscope, says Conley Rieder, a cell biologist who runs an NIH lab for biological microscopy at the Wadsworth Center in Albany, New York.

The NIH spent just over one per cent of its budget — \$150 million — on equipment in

1998. Just less than half of that was contained in R01 grants, averaging \$9,000 per grant.

A few programmes administered by the NIH and the NSF fund equipment that is beyond the scope of individual investigator grants, but scientists say these fail to meet the huge and growing need (see Table 1) and are impractical for buying equipment costing millions of dollars.

The programmes are growing more competitive as innovation draws scientists to demand the latest instruments. The NSF programme that funds shared instruments worth \$40,000 to \$400,000 saw applications jump by 30 per cent this year. And the NSF's major research instrumentation programme limits applications for off-the-shelf equipment to two per institution, says Nathaniel Pitts, who is in charge of the programme. Otherwise it would be "unmanageable", he says.

FASEB warned this year of a "broad degradation in the quality and availability of critical state-of-the-art research tools". It recommended that the NIH's \$35 million shared instruments programme should be boosted to \$80 million.

Senator Tom Harkin (Democrat, Iowa) has introduced a bill that would boost the NIH programme to \$100 million (see *Nature* 399, 621; 1999).

In the meantime, many investigators are using obsolete equipment. A 1996 report to the NIH's National Center for Research Resources showed that, in 1993, researchers



Brinkley (top) with the 1993 microscope for which the service contract has been cancelled.

Table 1 Major US grants available for off-the-shelf biological equipment

National Institutes of Health

Name of programme:	Shared Instrumentation Grant
Cost of instruments funded:	\$100,000–\$500,000
Minimum number of investigators:	3
Cost of programme:	\$35 million

National Science Foundation

Name of programme:	Major Research Instrumentation*
Cost of instruments funded:	\$143,000–\$2 million
Minimum number of investigators:	1 (but more preferred)
Cost of programme:	\$50 million (\$10 million for biology)

Name of programme:	Multi-user Biological Equipment
Cost of instruments funded:	\$40,000–\$400,000
Minimum number of investigators:	3
Cost of programme:	\$8.7 million

*Recipient institution must provide 30 per cent matching funds. Only two applications per institution per year are allowed.

were still using 70 per cent of the equipment from 1982 grants. Judith Vaitukaitis, the centre's director, says that investigators "will use anything they can get their hands on".

The problems with using dated equipment are more than just performance-related. Brinkley, a cell biologist at Baylor College of Medicine in Houston, Texas, says that Molecular Dynamics cancelled the service contract on his department's confocal microscope following the company's acquisition by Amersham Pharmacia Biotech in 1998. The microscope was bought in 1993 with grants from the NIH and NSF.

Because his research on breast cancer is "absolutely dependent" on the microscope, Brinkley may have to fork out tens of thousands of dollars for future repairs.

Alternative sources

Forced to innovate, investigators such as Opella are piecing together federal funds from disparate sources. The 900-MHz NMR spectrometer that Opella expects to receive this autumn was funded by six grants from three federal agencies.

But some have abandoned the quest for federal funds altogether. A consortium of nine leading New York institutions is financing its own NMR centre at the City University of New York in Manhattan, starting with five spectrometers. Helped by the city's business community, they have financed the project from their own funds and from private foundations and individuals.

Meredith Wadman

Astronomers fear 'devastating' cuts in US budget bill

Washington The American Astronomical Society (AAS) has branded as "shocking" proposed cuts in the budgets of the space agency NASA and the National Science Foundation (NSF). It says the cuts would "have a devastating impact on astronomical research".

The AAS statement was part of a growing drumbeat of criticism from scientific societies, aimed at spurring scientists to make their concerns known to members of Congress, before they return from their summer recess after the Labor Day holiday and vote on the bill funding the two agencies. The societies fear that complacency in the community may result in proposed cuts to research at the two agencies being retained.

In a bill funding NASA and NSF, passed by the House Appropriations Committee in July, the space agency's budget is cut by \$1 billion, to \$12.7 billion. Some \$240 million would be cut from space science and \$285 million from Earth science (see *Nature* **400**, 490; 1999). The NSF's budget is frozen at its 1999 level of \$3.65 billion.

In the AAS statement, Robert Gehr, the society's president, and Anneila Sargent,

president-elect, say they are "gravely disturbed" by these figures. "It is particularly shocking that the proposed cuts directly target NASA's Office of Space Science and NSF at a time when both agencies have so dramatically caught the public imagination," they add.

Fewer 'permanent' UK academic staff

London The proportion of staff on permanent contracts at UK universities is continuing to decline, according to new government figures.

The figures — compiled by the Higher Education Statistics Agency — show that 78 per cent of the 17,000 recruits at universities during the 1997–98 academic year were contract-based. The proportion of academic staff on permanent contracts is now 55 per cent, five per cent lower than three years ago.

The figures come at a time of continuing industrial action by the Association of University Teachers, the academics' main union, over pay and conditions. The rise of casualization is a major issue for the union.

Debate on stem-cell ethics develops

Washington The US government should fund research on human embryonic stem cells, say the American Association for the Advance-

ment of Science and the Institute for Civil Society — but it should not fund the process of deriving these cells from leftover embryos at fertility clinics.

Human embryos are destroyed in the process of taking the stem cells. The groups contend that human stem-cell research — including this process — can be conducted "in a fully ethical manner". But they say that public tax dollars should not be spent on derivation because it "raises ethical questions" for those who find intentional embryo destruction "morally wrong". This approach corresponds to that taken earlier by the Clinton administration.

The Coalition of Americans for Research Ethics, which opposes human embryonic stem-cell research, called the report's separation of use from derivation of the cells "ethically self-contradictory".

Geologists campaign to keep drill holes open

Munich German geologists are fighting to preserve two unique drilling holes in northern Bavaria, which are to be walled up next year according to German mining rules.

The holes, which are 4.5 and 9 kilometres deep, are the remnants of Germany's large continental drilling programme, completed three years ago. But scientists say the holes

are still of high scientific value in the investigation of tectonic tensions and fluid transport processes in the Earth's crystalline crust. The German science ministry has given a panel of scientists until the end of the year to work out a concept for future use of the drilling holes.

Meanwhile, Stanford-based geophysicist Mark Zoback, principal investigator of a planned international drilling project in the Californian San Andreas fault, has asked a team of German geologists to take part in the project (see page 5).

India plans to launch X-ray telescope

New Delhi India has announced that it will build a satellite-mounted X-ray telescope to be placed in orbit by 2004.

The Indian Space Research Organization, which has fabricated several remote-sensing and communications satellites, will design and build the new satellite at its centre in Bangalore. The X-ray detector will be built by scientists at the Tata Institute of Fundamental Research in Mumbai.

"Our detector will target X-rays in the 2–80 keV energy range which is not covered by other telescopes," says Ravi Manchanda, an X-ray astronomer at the institute. The satellite will also carry an instrument for

scanning ultraviolet radiation, being designed at the Indian Institute of Astrophysics in Bangalore.

Health minister to control GM regulation in Australia

Sydney The Australian government has decided to centralize regulation of genetically modified (GM) products in the hands of Michael Wooldridge, the minister for health, rather than in the science or environment ministries.

In a move seen as responding to public concerns about the health impact of GM products, Wooldridge will be given authority over their regulation until a permanent Office of the Gene Technology Regulator is established in 2001.

An office in Wooldridge's department will oversee five regulatory systems that separately control GM products in foods, therapeutic goods, agricultural and veterinary chemicals, industrial chemicals and imports/exports.

Research assessment under review in UK

London British scientists were this week asked to comment on the draft criteria and working methods that will be used to assess their

research in the year 2001. Views are being sought from higher-education institutions, professional bodies and other relevant organizations on the regular research review known as the Research Assessment Exercise.

A recently published document outlines the proposed working methods of the 62 panels dealing with different subject areas.

Those managing the exercise say this is the first time that such a consultation has taken place, and are hoping to use it as a way of increasing confidence in the procedure. The deadline for responses is 15 October.

Coining it in for disease research at Scripps

San Diego A highly prized collection of US coins worth an estimated \$5 million to \$6 million has been donated for auction to the Scripps Research Institute of La Jolla, California, to create an Institute for Childhood and Neglected Diseases.

The new institute will apply molecular biology techniques to research into and treat childhood diseases and disorders that afflict populations in developing countries.

Around 1,000 coins are to be auctioned in New York. They were donated by Rebecca and John Moores, the founder of a computer software firm from Texas and owner of the San Diego Padres baseball team.

Swift action needed to close the skills gap in bioinformatics

Sir — Bioinformatics, the manipulation and interpretation of the enormous volumes of data generated through research including genomics, proteomics, structural biology and dynamics of macromolecular assemblies, is a fast-growing area. It will lead to a better understanding of biology and of disease, and an improved ability to design targeted drugs. But the skills in computer science and information technology required are in such short supply that the pace of advancement of some biological and biomedical sciences may soon be limited.

To quantify this skills gap in the United Kingdom, the Biotechnology and Biological Sciences Research Council (BBSRC) undertook an evidence-based review. It concluded that, relative to supply, demand for bioinformaticians is high in academic institutions and in the bioindustries. An approximate 60% increase in staff numbers is needed immediately in the (non-medical) bioscience academic sector, and an approximate 45% increase is needed in large pharmaceutical companies. This situation will clearly become exacerbated with advances in genomics research and post-genomics.

Various training options, ranging from short courses to senior fellowships, need to be introduced or increased to address the skills gap. Diversity of options and of sponsors will ensure the best results. The BBSRC has increased by 60% the number of places sponsored on masters courses in bioinformatics and increased by a factor of two the number of PhD places. It is also introducing two bioinformatics summer schools for up to 60 staff and students a year.

The UK Medical Research Council (MRC) has also introduced training schemes in bioinformatics to support MSc and PhD studentships, as well as fellowships. In addition the UK research councils (BBSRC, MRC, Engineering and Physical Sciences Research Council, Natural Environment Research Council and Particle Physics and Astronomy Research Council) are to create at least two bioinformatics research groups, with training responsibilities, in computer science departments. Finally, the research councils are discussing future support with the Wellcome Trust.

The shortage of bioinformatics skills is not limited to the United Kingdom. The French 'génopôles', for example, will provide support for research and training in bioinformatics. Germany's research council, the Deutsche Forschungsgemeinschaft, has launched a programme to develop

research and training centres in bioinformatics. The Swiss Institute for Bioinformatics has recently opened to provide research and teaching opportunities. Japan has announced a five-year plan to double funds for genome-related research, so requiring significant bioinformatics expertise, and the Australian government plans to develop a national genomics research capability and bioinformatics networks.

The US National Institutes of Health (NIH) has yet to respond to the recommendation of the working group on biomedical computing for the establishment of up to 20 national programmes of excellence. But the National Institute of General Medical Sciences is expected to launch a programme in bioinformatics training in the next few months. Some NIH institutes already run interdisciplinary training programmes in genome analysis and interpretation.

The first international bioinformatics summer school has been launched at the University of Magdeburg, Germany, with industrial sponsorship from the Volkswagen foundation. The International Society for Computational Biology is obtaining industrial sponsorship for international training fellowships in bioinformatics. This programme will enhance the in-house training offered by some companies.

Clearly, there is progress in the provision of bioinformatics training. Equally clearly, however, a quick glance at the scientific press reveals that there is an increasingly large volume of data that can only be fully understood and used through the application of bioinformatics skills.

Marlie MacLean, Colin Miles

Biotechnology and Biological Sciences Research Council, Polaris House, North Star Avenue, Swindon, Wiltshire SN2 1UH, UK

France is still in the running on genomics

Sir — As one of the main coordinators of the genomics report published by the French Academy of Sciences, I must express disappointment about your recent News article (*Nature* **400**, 199; 1999). The title of the article, "France losing genome race", does not give a fair idea of the true spirit of our report.

The academy report attempts to give non-specialists an overview of the major current approaches and achievements in the field. It is true that the document underlines some weaknesses, gaps or shortcomings, for example in bioinformatics and in the development of start-up companies. But to state what could be improved by no means implies that the situation is hopeless. Our academy was asked to provide an independent vision of the state of genomics in France precisely in order to shed light on

strengths and weaknesses. On the whole, it is our conclusion that our scientists have not lost the race.

A few weeks after our report was published, the government responded by giving genomics a high scientific and strategic priority.

François Gros

Institut de France, Académie des Sciences, 23 quai de Conti, 75006 Paris, France

Japan's researchers could face more red tape

Sir — National research institutes in Japan will experience a major reorganization before becoming semi-autonomous agencies in 2001 (*Nature* **398**, 269; 1999). Although change may be necessary to improve efficiency, I do not have a rosy view of the reorganization. Scientists may have to struggle with even more red tape than at present.

This is what I predict will happen. Each ministry will try to retain power over the institutes that it controls at present. Preservation of power and budget is almost the basic principle of a government ministry. Existing institutes will be merged into a smaller number of research centres. It will be much easier for a ministry to control a smaller number of centres. But centralization may mean that individual institutes will lose their research independence. For scientists, centralization will certainly mean more tiers of administration above them. A research proposal must travel upwards through all the tiers of the hierarchy, winning approval at each stage. Having more tiers simply means more delay and more obstacles to clear. Politics and formalities rather than science may prevail in the decision-making process.

I believe that the hierarchical structure of the Japanese research system must be flattened to improve scientific performance and cut costs. Scientific performance will be the biggest concern for the new agencies, because it will be the criterion by which they are evaluated every three to five years, with the possibility that poor performers may be closed down.

The reorganization is planned as a means of cutting costs, and the agencies will be required to be much more efficient in spending money than the institutes are now. It is therefore crucial that the reorganization is done for the sake of performance and efficiency. If, instead, it is done for political reasons, the agencies will be laden with difficulties from the outset.

Kazuhiko Kobayashi

National Institute of Agro-Environmental Sciences, Ministry of Agriculture, Forestry and Fisheries, 3-1-1 Kannondai, Tsukuba, Ibaraki 305-8604, Japan

Seeing the wood for the trees

New journals are welcomed in rapidly changing disciplines, but — as Internet resources grow — many reviewers say we'd be better off saving the paper.

Journal of Optics A: Pure and Applied Optics

Editor-in-chief M. Bertolotti
Institute of Physics. 6/yr.
North America \$693, elsewhere £350

Journal of Optics B: Quantum and Semiclassical Optics

Editor-in-chief E. Arimondo
Institute of Physics. 6/yr.
North America \$573, elsewhere £289

Mark Saffman

While scientists love to argue, the proliferation of scientific journals is a topic that generates much consensus. The vast majority of my colleagues seem to agree that the number of journals is both too large and growing too rapidly. Besides the burdens that unchecked growth presents to economically pressed research libraries, the time spent by individual researchers in monitoring a large number of publications to keep abreast of developments in their field is of great concern.

A quick check of the list of journals with "optics" in the title in one form or another, available in print or electronically at my institution, resulted in 35 hits. It is therefore cause for some optimism when competing journals merge, with the goal of emphasizing quality, not quantity. *Journal of Optics A: Pure and Applied Optics* appeared in January 1999 as a merger of *Journal of Optics* (formerly *Nouvelle Revue d'Optique*) and *Pure and Applied Optics*. *Journal of Optics B: Quantum and Semiclassical Optics* is a relaunching of *Quantum and Semiclassical Optics* under a new title. *Journal of Optics* is now the official organ of the European Optical Society, published by the Institute of Physics in the UK.



DAVID NEWTON

It can be no secret that the intention is to launch a strong European competitor to *The Journal of the Optical Society of America*, A and B, which is second to none in terms of quality and prestige in the worldwide optics literature. Since optics research in Europe in no way lags behind that in North America, it is appropriate for the European optics community to publish its own flagship journal.

Although *Journal of Optics* has been published in its present form for less than a year, it is already a big success. It is attracting a large volume of high-quality work, although part B has had more success in securing long, fundamentally significant contributions than part A, which has an undue proportion of shorter, very specialized papers. Both parts are available in print and electronically.

In comparison with *The Journal of the Optical Society of America*, on the other hand, *Journal of Optics* lags significantly in securing truly important new research results. This is no surprise given the journal's tender age. If the new journals continue to emphasize high editorial standards, the American and European journals are likely to attain a more equal footing in the future. □

Mark Saffman is at the Risø National Laboratory, Postboks 49, DK-4000 Roskilde, Denmark.

By the people, for the people

New journals index

- | | |
|--|--|
| p11 Journal of Optics A: Pure and Applied Optics | p14 Current Opinion in Microbiology |
| p11 Journal of Optics B: Quantum and Semiclassical Optics | p15 CyberPsychology & Behavior |
| p11 Evolutionary Ecology Research | p15 Genetic Testing |
| p12 Animal Cognition | p16 Nutritional Neuroscience |
| p12 Chemoecology | p16 Public Health Nutrition |
| p12 Advances in Complex Systems | p16 BWP Update |
| p13 Pattern Analysis and Applications | p16 Mechanics of Time-Dependent Materials |
| p13 Inorganic Chemistry Communications | p17 Environmental Science & Policy |
| p14 Journal of Electroceramics | p17 Studies in History & Philosophy of Biological & Biomedical Sciences |
| p14 Journal of Seismology | p18 Current Opinion in Plant Biology |
| | p18 Granular Matter |

Evolutionary Ecology Research

Editor-in-chief Michael L. Rosenzweig
Evolutionary Ecology. 8/yr. \$272 (institutional, online), \$305 (online plus print); \$33 (individual, affiliated to an institution), \$125 (non-affiliated individual)

Peter McGregor

The publication of original research articles involves an uneasy alliance between academics and publishers, because the academics' aim of maximizing dissemination rarely equates with the publishers' aim of maximizing profits. But times change. Academics now have desktop technology to publish without the help of publishers. Add the distributing power of the Internet and conditions are ripe for the emergence of new journals such as *Evolutionary Ecology Research*.

This journal, owned and run by academics, aims to have liberal dissemination rules and to publish as inexpensively as possible.

These aims seem to have been met. Authors retain copyright and Internet access to articles is rather open. The annual library subscription includes hard copy (about 1,000 pages) and the Internet edition. This seems excellent value, given that the publishers' journal that spawned it, *Evolutionary Ecology*, costs libraries \$777. Curiously, individual subscription costs are similar for the two unless the person's institution subscribes.

But price isn't everything. Quality is hard to assess, but authors generally submit their best research to the journal with the highest citation rating and shortest submission-to-appearance time. These two features often replace cost as determinants of journals that, from an academic's perspective, must be in the library. It is not easy to predict how the citation rating and publication times of academics' journals will compete with those of journals backed by publishing conglomerates. As a consequence, many will closely monitor the co-evolution of *Evolutionary Ecology Research* and its competitors. □

Peter McGregor is in the Department of Animal Behaviour, Zoological Institute, University of Copenhagen, Tagensvej 16, 2200 Copenhagen N, Denmark.

Pulling the wings off flies

Animal Cognition

Chief editor Tatiana Czeschlik
Springer. 4/yr. \$324, DM496

T. S. Collett



Although its current practitioners might demur, animal cognition has a pedigree going back to at least the nineteenth century, when George Romanes collected

interesting, though perhaps not always reliable, accounts of the intelligent behaviour of members of the animal kingdom. To give an example from Romanes' *Animal Intelligence* (Kegan Paul, 1882): "I here found a robber wasp busied in lifting from the ground a large fly ... It had scarcely raised its prey a few inches above the ground when the wind caught the wings of the dead fly and they began to act like a sail. The wasp ... was blown ... in the direction of the wind, whereupon it let itself fall to the ground ... With eager industry [it] pulled off with its teeth the fly's wings which hindered it in its object. When this was quite done it seized the fly and flew off with it untroubled on its

journey through the air." Advances in ethology, psychology and neuroscience have since put animal cognition on a more rigorous footing, and have given its practitioners a new confidence.

With clever design, it is now possible to answer questions that a few years ago it would have seemed foolhardy even to ask. For instance, do monkeys that successfully use a tool have insight into how it functions? And cognitive concepts are becoming better grounded. Neuroscientists unravelling the mechanisms underlying ever more complex cognitive processes are finding seemingly comprehensible correlations between cognitive processes and neural circuits.

The subject is certainly worth a journal of its own. A laudable feature of *Animal Cognition* is its coverage of cognitive questions across a wide range of species. This should encourage cross-fertilization of work from very different intellectual traditions.

Perhaps the major problem for this new venture is the presence of several well-established journals which already publish papers in this field. How might an editor persuade authors to switch their allegiance to a new journal? First, offer rapid, rigorous and supportive reviewing, followed by speedy publication (*Animal Cognition* seems average). Low personal and library subscriptions (not obviously a Springer policy) will also increase a journal's readership and so its attraction to authors. Other ways of raising its profile are to have guest editors organize special issues, and to give the journal a tutorial as well as an archival function by inviting topical reviews (*Animal Cognition* encourages a nice mix of reviews and theoretical and empirical studies). The impressive and diverse editorial board augurs well for a forward-looking editorial policy. □

T. S. Collett is at the Sussex Centre for Neuroscience, School of Biological Sciences, University of Sussex, Brighton BN1 9QG, UK.

Smell, taste and other hot topics

Chemoecology

Editors-in-chief Désiré Dalose and Jacques M. Pasteels

Birkhauser. 4/yr. \$264, SFr368

(institutional); SFr158 (personal)

Tristram D. Wyatt

Frogs whose skin secretions are used to make poison arrows; chemical battles for niches on coral reefs; arms races by plants against herbivores and each other; the use of pheromones to attract mates — these are all aspects of chemical ecology, an important topic that is coming of age.

Researchers now have a much greater

understanding than in the past of the importance of chemically mediated interactions, including those of taste and smell, throughout the field of biology. This has happened in part because techniques of chemical analysis and synthesis have improved, but it is also because more biologists are now linking chemical ecology to interesting and more fundamental behavioural, ecological and evolutionary interactions.

Chemoecology aims to cover these interactions. It was relaunched in 1998, with new editors but an unbroken series of volume numbers. The first new volume covers a wide range of organisms, with a good balance between those of land and sea, but there is a preponderance of insects and plants (only one vertebrate paper out of 25). I imagine that this will change over future issues. The journal publishes a relatively small number of pages (200) per year, compared with about 10 times more than this published by the other main journal in the field, *Journal of Chemical Ecology*.

There is room for two journals in the field as it continues to grow. *Chemoecology* could have its greatest influence if it can increase the number of mini-reviews it carries, particularly if these integrate studies of diverse taxa, vertebrate and invertebrate and if, as the editors intend, the journal becomes a forum for current debates.

Chemically mediated systems can offer unique opportunities for researchers to investigate hot evolutionary topics, for example in sexual selection and the evolution of mate communication.

Paradoxically, as chemical ecology becomes more mainstream, the challenge for journals like *Chemoecology* is to attract authors from higher-impact evolution and ecology journals. An important factor will be the likelihood of articles being found by online electronic searches. *Chemoecology* appears in many abstracting services but is not yet in *Current Contents*, so in this area it still has some way to go. □

Tristram D. Wyatt is in the Departments of Zoology and Continuing Education, University of Oxford, Oxford OX1 3PS, UK.

Complex issues in a specialist field

Advances in Complex Systems

Editor-in-chief Eric Bonabeau

Editions Hermès. 4/yr. Europe FF1,050, elsewhere FF1,250 (institutional); Europe FF682.50, elsewhere FF812.50 (personal)

H. Eugene Stanley

Another new journal? When virtually every paper is now deposited in preprint form on servers that are accessible by everyone: rich

and poor, old and young, beginning student or grand old master? Who, then, is expected to shell out for four issues a year when virtually all journals are free, or essentially free, in electronic form? Who has the optimism to believe that authors will choose to publish in a journal that may have no free electronic distribution, no listing in Science Citation Index or Web of Science, and no fewer than three months' delay between issues? In short, how can any new journal survive?

It is thus truly remarkable that the editors of *Advances in Complex Systems* have undertaken to initiate a new journal in a field in which there are already a number of journals publishing papers. This journal corrects the unfortunate situation faced in many fields that are truly multidisciplinary. Papers published by, say, a physicist in a physics journal are unlikely to be found by, say, a psychologist who normally does not scan the pages of physics journals — except by using Web of Science or other electronic media that can be searched by keyword.

The list of editorial board members, as well as the authors of the papers appearing in the first few issues, is truly outstanding, with computational scientists, biologists, physicists, chemists, an economist, a political scientist, a psychologist, a cognitive scientist and even a nuclear engineer thrown in for good measure! The journal's editor-in-chief, Eric Bonabeau, heads the Santa Fe Institute, which has made its mark on world science by creating a unique environment where scholars can meet to think and discuss in an idyllic setting, isolated both culturally and geographically from the competitive hustle of traditional science research centres that have to obtain funding for specific projects.

The bottom-line question is whether this journal is a worthwhile addition to a typical library which may be in the throes of reducing, not augmenting, its list of journal

subscriptions. The quick answer is "yes", at least for a year, to see if there is as much interest among the library users as there is among this reviewer and his colleagues. □

Eugene Stanley is at the Center for Polymer Studies, Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, Mass 02215, USA.

Quick work but variable quality

Pattern Analysis and Applications

Editor-in-chief S. Singh
Springer. 4/yr. £156 plus postage

Martin A. Giese

Pattern recognition has been under investigation for more than 30 years. This interdisciplinary subject has a lot of interesting technical applications, for example in image processing, robotics and multimedia systems. A number of different journals cover pattern recognition in general and in the context of different specialized disciplines, such as computer vision, neural networks, cognitive science and machine learning.

Within this growing body of literature, the new journal *Pattern Analysis and Applications* tries to focus in particular on novel techniques and industrial applications of pattern recognition. Additionally, it wants to offer a forum for the publication of benchmark studies and comparisons of different methods. Each issue contains four to six original articles (without length constraints), one or two book reviews and a calendar of conferences and workshops on related topics. Abstracts are accessible through the Internet.

The issues so far have provided a good balance of contributions from laboratories in different nations, and between articles

about theory and about applications. A number of articles that compared and evaluated different methods seemed to me to be particularly useful. With turn-around times between four and seven months, it is definitively among the fastest-publishing journals in the field.

The quality of the contributed articles has been variable but reasonable — though it seems that many authors do not contribute their most influential work. *Pattern Analysis and Applications* will only be able to fulfil its aim to become one of the leading journals in the field if more authors are willing to provide articles of central general relevance. A subscription can be recommended, in particular for institutions that are working on applied aspects of pattern recognition. □

Martin A. Giese is at the Center for Biological and Computational Learning, Massachusetts Institute of Technology, 45 Carlton Street, Cambridge, Mass 02139, USA.

New developments in an old discipline

Inorganic Chemistry Communications

Editors U. Belluco, A. S. C. Chan, D. G. Nocera

Elsevier Science. 12/yr. DfI 705, \$358 (institutional); DfI 212, \$107 (personal)

Dante Gatteschi

Inorganic chemistry is a very old discipline. One of humanity's greatest leaps forward corresponds to the early implementation of chemical techniques for the treatment of metals and ceramics. In recent years, inorganic chemistry has renewed itself by exploring its frontiers with biology, materials science and environmental science.

Metallo-proteins and metallo-enzymes, metal ions in medicine, ceramic superconductors and organometallic precursors for semiconductors, the role of heavy metals in the environment — these are just some of the many different themes that are of interest to inorganic chemistry. At the same time the traditional areas of catalysis and reactivity have been extended, taking advantage of the most recent experimental and theoretical techniques for the investigation of molecules and their interactions. New synthetic routes for traditional compounds have been developed that are more environmentally friendly. Furthermore, the inorganic contribution to the development of supramolecular chemistry cannot be overlooked.

The monthly periodical *Inorganic Chemistry Communications*, first published in January 1998, complements journals such as *Inorganica Chimica Acta*, *Polyhedron* and *Journal of Organometallic Chemistry*. It is



DAVID NEWTON

dedicated to the rapid publication of important reports or original advances in inorganic and organometallic chemistry: in its first year about 130 papers were published within three months of the submission date.

Inorganic Chemistry Communications is also available online through the ScienceDirect programme. This journal, though pricey, is of a comparable standard to others in the field. With its rapid publication of new manuscripts, there is no doubt that it is a useful tool for all researchers who want to keep up to date with the latest developments in an old discipline. □

Dante Gatteschi is in the Department of Chemistry, University of Florence, 77 via Maragliano, Florence, Italy.

Materials with added value

Journal of Electroceramics

Editor-in-chief Harry L. Tuller
Kluwer. 4/yr. \$255, Dfl510 (institutional); \$105, Dfl198 (personal)

Robert W. Cahn

A number of established broad-spectrum ceramics journals already exist — such as *Journal of the American Ceramic Society* and *Journal of the European Ceramic Society* — as well as some broader-spectrum materials journals such as *Journal of Materials Science*. They all include papers concerned with such materials as ferroelectrics, magnetic ferrites, solid electrolytes, piezoelectrics and semiconducting and superconducting oxides. According to the editor's preface in the first issue of the new journal, this group of ceramics is collectively termed 'electroceramics'; he judges that the literature related to these materials is too widely scattered, so that the various research communities involved remain unaware of much of it. This is the declared motive for inaugurating *Journal of Electroceramics*.

The journal shows every sign of covering the full range of electroceramics. Indeed, it goes a little beyond: one issue is devoted to sensors, and includes a paper on giant magnetoresistance, a property of metallic multilayers. Another special issue is most interestingly devoted to "nanostructured materials for energy applications", and includes a paper on nanostructured catalysts — again, an extension of the electroceramics concept. I point out these innovations in admiration, not criticism. 'Mainline' papers include several on processing innovations, non-stoichiometry and associated defects, impedance spectrometry, ionic conductors and ceramic superconductors.

A point of concern is the institutional subscription price. It is usual for publishers

to tempt little fishes in with gently smiling jaws, and for their crocodilian propensities to be manifested only later. Kluwer, it seems, is revealing itself as a crocodile from the word go. Nevertheless, the new journal warrants a place in those technical libraries that can afford it and whose readers include a posse of electroceramicists. □

Robert W. Cahn is in the Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK.

Earthquakes — the grand and the gritty

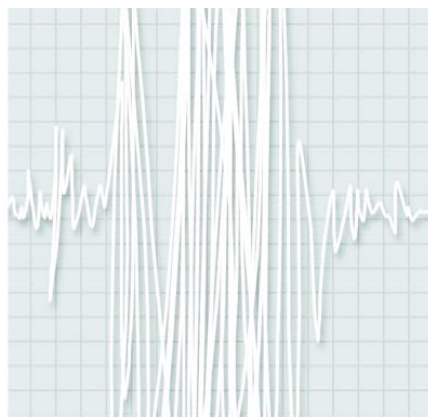
Journal of Seismology

Editor-in-chief A. Udías
Kluwer. 4/yr. \$265, Dfl530 (institutional); \$99, Dfl150 (personal)

George Helffrich

As a discipline, seismology is both grand and gritty. The gritty part is where seismology happens — the rock, soil, mud and water that moves during and between earthquakes. The grand part is the underlying theory and the reward gained by applying that theory in order to understand the behaviour of inaccessible parts of the Earth. While there are controversies, the drive uniting seismologists to increase the general understanding of the Earth comes from the communities dealing with seismic hazard and with seismic source discrimination. Seismology journals publish an intriguing blend of practice, theory and planetary observation that informs a diverse community united by a need to know.

On account of its diversity, I think the discipline is best served by placing seismology in the broader context of Earth science (or planetology/astrophysics, to include helioseismology). The two-year-old *Journal of Seismology* takes the opposite view, focusing on the subdiscipline rather than the context. The most informative elements in my present seismological reading list are broadly constituted geophysical journals, plus



Science and Nature, but the list also includes some specialized seismological society-sponsored journals such as *Bulletin of the Seismological Society of America* and *Seismological Research Letters*.

The list is crowded, and the *Journal of Seismology* competes for an increasingly pressed personal subscription budget. Its institutional subscription rate exacerbates the distressingly noxious negative-sum game involving my university's library budget too much to permit me to warmly commend it to my colleagues. □

George Helffrich is in the Department of Earth Sciences, University of Bristol, Queens Road, Bristol BS8 1RJ, UK.

Bugs, glorious bugs

Current Opinion in Microbiology

Editors Julian Davies and Stanley Falkow
Current Biology. 6/yr. £500, \$822 (institutional, print only); £125, \$206 (personal, print only); £150, \$247 (personal, print plus online)

Peter H. Williams

There has rarely, if ever, been a more exciting time to be a microbiologist. Microbes do such interesting things (metabolically speaking), live in such interesting places, respond to their environments, communicate with one another — and 99% of microbes haven't even been discovered yet. Gone is the complacency of the antibiotic era. Infectious diseases continue to wreak havoc in the poorer regions of the world, while in the West tuberculosis is back and 'new' infections drive us to panic. Microbes also play crucially important roles in the manufacture of food and drink and in the pharmaceutical, agrochemical and biotechnology industries.

But perhaps most exciting is that microbiology is in the grip of a technological revolution that is fundamentally changing how we study and exploit microbes. Genomics, proteomics, DNA arrays and the associated developments in bioinformatics have ignited an information explosion that I am sure (and this is the *really* exciting bit) will raise at least a hundred times as many new questions as it answers.

How will microbiologists get their heads round all the new data, make sense of flawed hypotheses, come down off hobby-horses, release bees from their bonnets? I recommend they try *Current Opinion in Microbiology*. Eight major aspects of microbiology are covered each year in six issues: host-microbe interactions in bacteria; cell regulation; ecology and industrial microbiology; techniques; host-microbe interactions in fungi/viruses/parasites; antimicrobials; genomics; and growth and development. Each aspect

comprises an editorial overview of up to 16 reviews by recognized experts in the field, combining timely assessments of accumulating data with occasionally controversial but always interesting points of view.

The papers are generally of high quality, well written, beautifully illustrated and fully referenced (with, usefully, brief résumés of key features of a few of the more important citations). The journal is not, of course, a substitute for actually going and reading the primary data in the original papers. But it does give the overworked, overwhelmed microbiologist an invaluable, authoritative and convenient way of keeping up to date with a familiar field or of beginning to come to grips with an unfamiliar one. And by revisiting each aspect annually, the editors have ensured a regular 'state of the nation' appraisal that will help microbiologists make sense of the information explosion. □

Peter H. Williams is in the Department of Microbiology and Immunology, University of Leicester, Leicester LE1 9HN, UK.

Computers and the human brain

CyberPsychology & Behavior

Editor Mark D. Wiederhold

Mary Ann Liebert. 6/yr. USA \$143, elsewhere \$182 (institutional); USA \$79, elsewhere \$119 (personal)

Amanda Parker

Separating the interesting issues related to the impact of the Internet, multimedia and virtual reality (VR) from their respective media hype is never going to be easy. Like these products of the global information age, scientific research into their effects on behaviour and society is likely to be both prolific and of a very variable quality. The challenge for a journal devoted to this area, then, is likely to be one of selection.

The range of articles in *CyberPsychology & Behavior* (C&B) is wide. The journal is primarily aimed at healthcare providers who are interested in the use of advanced technology to reduce the cost of healthcare. Consequently, many articles discuss advances in virtual reality that allow its application in a mental health setting. For example, patients with acquired brain injuries, neurological disorders and developmental disabilities may benefit from VR training in a range of simulated environments. VR also has exciting possibilities in psychotherapy. Phobias may be treated with more precision in a virtual environment which can be switched off when necessary: recently published C&B articles show that fear of flying is one area in which progress is being made. The benefits of VR in treatment of post-traumatic stress



disorder have also been recently discussed.

On the other hand, evidence of the Internet's value in psychotherapy is less convincing, and articles presenting research into the Internet's impact on behaviour tend to rely on opportunity sampling and number crunching to reach their conclusions. These papers take a very different approach from the patient studies discussed above, which makes for an uneasy balance in the journal. *CyberPsychology & Behavior* provides a forum for a range of topics that may turn out to be too wide for comfort. □

Amanda Parker is at the School of Psychology, University of Nottingham, University Park, Nottingham NG7 2RD, UK.

Testing the market in search of a niche

Genetic Testing

Editor Fred Gilbert

Mary Ann Liebert. 4/yr. USA \$186, elsewhere \$238 (institutional). USA \$99, elsewhere \$150 (personal). USA \$50, elsewhere \$75 (student)

Peter Harper

Although I had seen the initial publicity for this new journal, it was interesting to see how the first 18 months of production has matched up to the initial stated aims. The journal has a distinguished molecular geneticist as editor-in-chief and a broad-ranging, well-respected editorial board (though it is not clear whether they are functional rather than decorative).

The editorial in the first issue makes it clear that *Genetic Testing* is intended to have a wide readership including the professional genetic testing community, those who develop

or perform such testing, and those "whose expertise and interests fall within the scope of utilization, interpretation, and delivery of genetic testing." The journal invites original papers on diagnostic genetic testing of all kinds, newborn and carrier screening, technical approaches to testing and ethical and social aspects.

The balance of papers is very much towards reviews, with several collections of papers resulting from symposia occupying most of the space. The first of these, spread over two issues, is a series of reviews on the frequency of genetic disorders and mutations in Jewish populations, concentrating on disorders of specific significance in Ashkenazi Jews. This is a good example of material that is often scattered too widely or not published at all, so it is good to see the collection appear here, though one might ask whether it would not find a wider readership in a small monograph.

The second collection occupying an entire issue relates to genetic testing and Alzheimer's disease. Again the broad conclusions had appeared earlier (in *Nature Medicine* 7, 757-759; 1998) but there is much detail in addition that deserves publication.

The third collective series is on the topic of genetic privacy. I found this less valuable, perhaps largely because many contributions, particularly those on insurance and genetic tests, were parochial and written in the specific context of the United States rather than internationally.

In terms of original articles, I do not think that any of the papers I saw rate highly, nor would I consider the technical reports particularly important. There are some disease maps of the human genome which seem rather redundant now that databases cover this information and are regularly updated.

It would be nice to think that scientists, genetic counsellors and ethicists as well as

others might all be readers and gain from the interactions, but I rather doubt whether this will actually happen. I suspect that this may become a journal more related to clinical practice and its social and ethical implications than to high-profile original contributions. Though this may not have been the original intention behind the journal, it could turn out to be the right course; perhaps it indicates a niche that has not been filled by other journals so far.

My advice to the editors would be to encourage the journal to develop along whichever direction it looks like taking and not be too worried if it fails to attract high-profile original work. There should be plenty of opportunities for useful contributions over the next few years. □

Peter Harper is at the Institute of Medical Genetics, University of Wales College of Medicine, Heath Park, Cardiff CF4 4XN, UK.

Evolving discipline finds a forum

Nutritional Neuroscience

Editor Chandan Prasad
Harwood Academic. 6/yr. 86 euros

Public Health Nutrition

Editor Barrie Margetts
CABI. 4/yr. £175, \$310 (institutional); £48, \$80 (special rate for members of The Nutrition Society)

Gema Frühbeck

Given the cornucopia of nutrition literature, one could justifiably ask if we need yet another journal in this field. Nutrition in the twenty-first century will be based increasingly on creating an interdisciplinary field of epidemiology, structural biology, integrative physiology and molecular pathogenesis. This should ensure that we revolutionize the ways in which we investigate, diagnose and treat nutrition-related health problems.

The late twentieth century has seen a number of fundamental transformations in the study of the brain. Recently, the qualitative approach crucial to understanding nutrition has been embraced by neuroscientists to the extent that the subject may be on the verge of a golden age, in which fields once thought impenetrable can be explored in detail. The publishing venture of *Nutritional Neuroscience* is an acknowledgement of the growing community of researchers dedicated to studying these very exciting topics from a wide range of disciplinary perspectives. The enterprise will play an important part in bridging the gap between neuroscientists

and nutritionists by focusing on eating disorders, at the same time as trying to unravel the complex underlying neuroanatomical and pathophysiological processes.

Given the Jeremiahs who prophesy that it is foolish to launch journals in such difficult financial times, the publishing venture of two new nutrition-oriented journals must be congratulated.

Nutritional Neuroscience is a handsome periodical covering an admirably broad range of topics including the effects of the various components of diet, dietary supplements and food additives on neurochemistry, neurobiology and behavioural biology. The neural and hormonal control of food intake, and dietary considerations in the management of neurologic and psychiatric disorders are also covered. It is the first journal to focus specifically on this area, and fills an important niche. The journal therefore merits reading by neuroscientists, nutritionists, neurologists and psychiatrists.

Public Health Nutrition was launched in March last year to cover studies involving nutritional epidemiology, nutrition-related health promotion, evaluation of the effectiveness of intervention studies aimed at improving health, the role of nutrition in high-risk and vulnerable groups as well as population-based research related to primary prevention of illness. This periodical will be valued by nutritionists and dietitians involved in nutritional epidemiology research, primary prevention and public health, together with epidemiologists and health-promotion specialists who are interested in the increasingly recognized role played by nutrition in disease prevention.

This attractive journal has an outstanding editorial board, with members representing 11 countries and committed to supporting, developing and maintaining an international perspective. So far, about half of the articles have come from Europe, the United States and Canada. Interestingly, 39 per cent of articles have been submitted by groups from different countries working in collaboration.

The period from submission to acceptance is generally two to three months, although the lag for actual publication is around five to six months. *Public Health Nutrition* fulfils

the requisites of a high-quality journal and is therefore most welcome.

Web-surfing scientists may also enjoy visiting the electronic address of The Nutrition Society (<http://www.nutsoc.org.uk/phnj>) to browse through the papers of all of the issues of *Public Health Nutrition*. Updated regularly, the site is extremely user-friendly. □

Gema Frühbeck is in the Department of Endocrinology, Clínica Universitaria de Navarra, University of Navarra, 31008 Pamplona, Spain.



Birds unlimited

BWP Update: The Journal of Birds of the Western Palearctic

Editor Malcolm Ogilvie
Oxford University Press. 3/yr. Print only. £85, \$150 (institutional); £45, \$75 (personal)

Juha Merilä

Ornithology is in a unique position among the biological sciences: practitioners are, with ever-increasing pace and in meticulous detail, being deluged with new literature about the lives and appearances of their study objects. Caught in this explosion of information, the relatively recent (1977–94), nine-volume, 7021-page reference work, *Handbook of the Birds of Europe, the Middle East and North Africa* (Oxford University Press, £795), also known as *Birds of the Western Palearctic* (BWP), is already outdated in many places. A new journal, *BWP Update*, was therefore launched to keep this ambitious and authoritative work as comprehensive as possible.

Since its launch in April 1997, *BWP Update* has covered 30 species (out of the 797 in BWP), including short descriptions for 10 species new to the Western Palearctic. The style and format of the book are largely maintained, but a new section on conservation has been added. Distribution maps are now re-drawn and updated in colour.

With its lavish and well-edited appearance, *BWP Update* may appeal to both professional and amateur ornithologists and birdwatchers, although many may be put off by its high subscription price — especially since much of the 'new information', such as updates on population trends and distribution ranges, has already appeared in better formats in recent books. On reflection, an easily updateable and searchable online/CD-ROM edition of *BWP Update* would be a more valuable innovation than this journal for those using BWP on a regular basis. □

Juha Merilä is in the Department of Population Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18 d, SE-752 36, Uppsala, Sweden.

Depends how much time you've got

Mechanics of Time-Dependent Materials

Editors Wolfgang Knauss and Igor Emri
Kluwer. 4/yr. Dfl525, \$262

David P. Pope

It seems only yesterday that we were all decrying the proliferation of journals (the comment by one of my colleagues some



years ago that there are so many journals sprouting up that soon there will be a *Journal of Shear Stress* comes to mind), while now I mostly receive announcements that our library is cancelling journals. Yet new journals do continue to appear—and we wonder in each case if the world would have been better off with the trees.

There is no doubt that some new journals appear because a particular field is currently 'hot', only for the journals to disappear after the field cools. But in other cases, there is a community of scholars working on similar problems but in different contexts, and doing their research in different departments and fields. The logical next step is to establish a journal to connect these scholars and to provide a forum for their output. How do we then decide if the tree/journal balance is a favourable one?

I believe that the decision depends on the generality of the subject, its long-term importance to science and technology in general, and how large the potential community is. In the case of the potential audience for *Mechanics of Time-Dependent Materials*, there is a large and disparate community working in the field, but they are separated by material systems, discipline and department. As a result, this critical materials problem tends to be marginalized in the typical journal. Many of us study the microstructural details that lead to a time-dependence in a given material behaviour. But we tend not to study the details of the time-dependence itself, and we are even less likely to model it carefully so that it can be described in analytical detail.

The goal of this journal is to provide a forum for researchers whose focus is on the time-dependence itself and to inject more science into issues of durability and ageing than has been the case in the past.

The journal has clearly made a slow start. Five issues have appeared since its inception in 1997, with the sixth due out soon, but the frequency of publication is increasing. The articles published are true to the journal's goals, focusing on time-dependence itself in materials ranging from foams to epoxies to polycrystalline metals to composites. Those for whom

the details of time-dependent material behaviour is essential should read and publish in this journal. □

David P. Pope is in the Department of Materials Science and Engineering, University of Pennsylvania, LRSB Building, 3231 Walnut Street, Philadelphia, PA 19104, USA.

Working together for the Earth

Environmental Science & Policy

Editor-in-chief Joe Wisniewski
Elsevier. 6/yr. \$477, DfL 940

David G. Victor

The objective of this journal is to "build bridges among and between scientists, policy makers and decision-makers in industry". In practice, as this journal shows, that worthy goal is difficult to meet.

Of the journal's strongest articles, most are written for specialists and build few bridges with other disciplines or decision-makers. Policy implications are left latent or are stated so broadly as to be of little use for real decision-making. Such studies belong in traditional disciplinary journals. Glaring in its absence from the four numbers provided for this review was the discipline of economics, which is a vital guide as policy makers struggle to allocate scarce resources. Also scant is attention to the implications of uncertainty for policy choices.

About half the articles reviewed are brief surveys or comments on existing research. They lack the depth needed to push the frontier of policy research or to synthesize research across disciplines to create new insights and bridges. Perhaps such articles would appeal to readers who have a broad interest in policy and environmental science, but few general readers are likely to glance at this journal—not least because the publisher has put the cost of subscriptions in the stratosphere.

However, a few gems shine. Among them is the study of carbon that is accumulating in the Canadian forest products sector (FPS) such as in wooden buildings, landfills, pulp and paper (vol. 2, no. 1, 25–41). Most forest carbon is in the forests themselves, but the authors show that FPS carbon is a large part of the net flux of Canadian forest carbon and a growing fraction of the FPS flux crosses political borders because forest products are increasingly traded overseas.

Such studies underscore how tricky it will be to create an accurate and comprehensive system for tracking carbon. This year, *Environmental Science & Policy* devoted a whole issue to this problem. The Kyoto Protocol, which limits fluxes of carbon dioxide and other gases that cause global warming,

makes the issue urgent. The papers identify the technical hurdles and show that countries will be unable to assure their compliance with the protocol until an accounting system is agreed.

This journal has not yet found its voice, but its goals are important and there are hints as to how they can be met. □

David G. Victor is at the Council on Foreign Relations, 58 East 68th Street, New York, NY 10021, USA.

Modern questions in a traditional form

Studies in History and Philosophy of Biological and Biomedical Sciences

Editor N. Jardine

Pergamon. 4/yr. \$197, 176.52 euros (institutional); \$45, 40.39 euros (personal)

W. F. Bynum

Few disciplines seem to have such a high ratio of journals to practitioners as the history and philosophy of science, medicine and technology. The history and the philosophy of science have tended to drift apart over the past few decades, because of increasing specialization and the discovery of the subject by scholars originally trained in history or philosophy. For some, the social history of medicine or science is pretty much all there is to it. Traditional philosophers of science are bemused by the emphasis on the social, as opposed to the cognitive, content of this historiography, and feel beleaguered by the burgeoning field of bioethics.

Studies in History and Philosophy of Biological and Biomedical Sciences, the product of a 1998 fission of an existing journal (*Studies in History and Philosophy of Science*), combines both traditional and cutting-edge features. It is traditional in offering space for both philosophers and historians; traditional, too, in that its philosophical balance is within the realism with which most scientists identify (discovery rather than social construction is the name of the game). Thus, Nils Roll-Hansen has some sharp criticisms of Bruno Latour's anthropological account of laboratory life, and Martha Keyes' analysis of the relationship between prions and molecular biology's central dogma takes the science seriously.

At the same time, the articles in the first four issues were written by professionals within the broadly defined field of 'science studies'. They are aware of the variety of approaches that have been used to examine the creation of scientific knowledge. They articulate the wider issues of social responsibility and power that permeate the modern scientific enterprise. Almost all of the articles

deal with nineteenth- or twentieth-century science, and most conform to the wish, expressed in the inaugural editorial, of providing information and interpretation that should interest practising scientists. Topics include interferon, *Caenorhabditis elegans*, gene mapping, *Helicobacter pylori* and the images of Barbara McClintock.

Consequently, this new journal can be welcomed as an important forum for contemporary debate. Within its first two volumes it has grown from two to four issues a year, in itself perhaps a reflection that the field is still growing. □

W. F. Bynum is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BE, UK.

New shoots to replace the old?

Current Opinion in Plant Biology

Editors Keith Roberts and Chris Somerville
Current Biology Publications. 6/yr. Print £500, \$822 (institutional); £125, \$206 (personal). Online \$227 (personal only). Print plus online, £150, \$247 (personal); £60, \$90 (student)

Ortrun Mittelsten Scheid

Current Opinion is the family name of quite a prolific clan. Last year, the parents at Current Biology Publications announced the birth of a new child, *Plant Biology*. Its bright green colour distinguishes it from its brothers and sisters on the shelf, but once opened it cannot conceal extensive family likenesses.

The concept of a bimonthly collection of invited, well-prepared reviews has proven very successful in other rapidly developing fields such as biotechnology, genetics and development, cell biology, immunology,

neurobiology and others. But unavoidable overlap with other family members and with other review journals (related by the marriage between CBP and Elsevier) suggests a birth-control plan would now be in order.

The issues are grouped around growth and development (February), genome studies and molecular genetics (April), physiology and metabolism (June), plant microbe interaction (August), cell signalling and gene regulation (October), and cell biology (December). Each issue starts with Paper Alert, a personal but widely agreeable selection of important primary publications in all six fields, and Web Alert suggesting useful Internet sites. The editorial overview that follows usually comes closest to the family name and the intention expressed in the first volume — “to provide informed opinions about the burden of evidence and possible directions for new discoveries”.

The reviews I read had much less of the opinion component. However, they were excellent summaries with informative introductions and a good compilation of basic as well as the most recent data, supported by suitable figures when necessary, and were followed by an extensive and ranked reference list. So far little use has been made of the ability to offer supplementary material such as video documentation via the publisher's home page, but this may change in the future. The concept of covering the same areas in the same month each year will render the journal an attractive source of information, also for readers who are not fond of systematic or computer-based literature search.

The price is high, but not unreasonable. Some institutions will not be able to add it to the budget without cancelling other subscriptions. But there is a good chance that this new member of the family will develop well, protected by its caring parents, the publishers, and its prominent godfathers

Roberts and Somerville. It is worth investing in its future, and going for a change of generation on the library shelves. □

Ortrun Mittelsten Scheid is at the Friedrich Miescher Institute, P.O. Box 2543, CH 4002 Basel, Switzerland.

Grains of wisdom unite two disciplines

Granular Matter

Editors-in-chief R. P. Behringer and H. J. Herrmann
Springer. 4/yr. \$332, DM498

Dov Levine

Granular materials such as sand are agglomerations of many macroscopic, hard, heavy particles, or grains. The grains don't deform appreciably, there are no randomizing effects of temperature and they dissipate energy in collisions and through friction.

Although granular materials may appear simple, the study of their physical properties is difficult: they are quintessential many-body systems, and knowing the interaction between two grains does not imply understanding of the system as a whole. They exhibit a wealth of fascinating behaviour, including novel pattern formation and interesting flow characteristics. Because of their strong dissipation, granular materials jam into static states whose properties are not amenable to standard statistical mechanics. The way force propagates in these states is different from any other system.

Engineers have long studied granular materials because of their importance in industry, while physicists have become interested more recently. The questions asked by the two communities are not always different, but the background assumed, the experimental tools used and the notation employed certainly make the papers look different. And, as is perhaps common in active interdisciplinary fields, there is a certain tension between the camps and the search for common ground is not always smooth.

Granular Matter was created to help find that common ground. Since there are so many journals in physics and engineering already, the editors' foremost aim is unification. This aims to be the journal of choice for the entire community, so that researchers of all backgrounds will be made aware of different approaches. This is a commendable idea for this exciting field. The first issues have had a nice collection of papers with a varied mix of authors, so *Granular Matter* is off to a good start. It is to be hoped that physics and engineering libraries alike will choose to add it to their shelves. □

Dov Levine is in the Department of Physics, Technion, 32000 Haifa, Israel.



DAVID NEWTON

Last days in Arcadia

The genetic code and Francis Crick shared a memorable birthday party.

John Cairns

The entire genetic code had been worked out by 1966. In a sense, this was not really a discovery. Obviously there had to be some kind of code and, in the end, it was deciphered more by brute force than subtle argument, using short molecules of RNA to drive protein synthesis in a test-tube. But the table showing the meaning of each of the 64 possible triplets was the culmination of all that had gone before.

In the 13 years since Watson and Crick had revealed their model for the structure of DNA, the essential ingredients that handle biological information had been identified — DNA sequence, operons, repressors, messenger RNA, transfer RNA and ribosomes, plus the various classes of enzymes needed to make those ingredients. These were formidable discoveries, and they changed the science of molecular biology from an esoteric field inhabited by a small coterie into the discipline that has dominated biology ever since.

In those days, there were far fewer scientists. Conferences did not tread hard upon each others' heels, and a single symposium at the Cold Spring Harbor Laboratory in New York state could just about cover all of molecular biology. The first breach into the code had come with the discovery that poly-U — an RNA chain of uracil nucleotides — translated into a chain of the amino acid phenylalanine. This, incidentally, contributed to the feeling of many molecular biologists that God was on their side, because poly-U was the easiest RNA to make and poly-phenylalanine is barely soluble and therefore easily detected. The rest followed soon after, making the complete deciphering of the genetic code the natural subject for the 1966 symposium.

It is the custom, halfway through each symposium, to hold a banquet preceded by a cocktail party — something is needed to relieve the fatigue and tensions brought on by listening to 15 or 20 talks each day. By chance, the 1966 symposium, celebrating the birth of the genetic code, coincided with Francis Crick's fiftieth birthday. To mark the occasion, Jim Watson devised a Bacchanalian interlude to take place during the cocktail party. I shall cast a veil over the details. But it was sufficiently lurid that Max Delbrück, hearing about it — and being himself something of an expert in embarrassing practical jokes — insisted that there should on no account be any party for his sixtieth birthday, due to coincide with his teaching that summer's Cold Spring Harbor phycomycetes course.

So our scene is set. The party is under way. Crick survives Watson's interlude. A telegram from England is read out, mysteriously signed "Elizabeth", which congratulates Francis on having managed to reach the age of 60 (*sic*). Rollin Hotchkiss reads a narrative poem, composed for the occasion and entitled "A Happy Crickmass", a tale of arms and the man which seems to go on almost as long as Virgil's epic. And the visitors mill around on the sloping sunlit lawn between the dining hall and the harbour, in a vinous stupor, secure in the knowledge that there are no lectures on the evening of the banquet.

At the time, my view of the proceedings was overcast by the gloomy pragmatism that affects most administrators, but any visitor from outer space would surely have judged this as an Arcadia peopled by innocents. The territory between biochemistry and genetics, long virtually uninhabited, was now open to all. Furthermore, thanks to Sputnik One, there was ample support for science, and

funding agencies had not yet fulfilled Leo Szilard's prophecy that, with more money to spend, they would become fussy and insecure. Those were golden days indeed.

A third of a century later, molecular biology has lost its air of relaxed innocence. In the 1960s, it was primarily the study of information stored in one dimension — the DNA molecule. The prevailing view was that amino-acid sequence was sufficient to determine the three-dimensional structure of a protein; this, therefore, was not a pressing problem. The only structure of interest was the double helix and, as Alfred Hershey once said, it was remarkable in being the only instance where structure had anything useful to say about mechanism. So the era of one-dimensional molecular biology in the mid-1960s was a time of maximum simplicity.

Since then, however, we have become able to find out how things work in three dimensions, and we are seeing that billions of years of evolution have allowed (or forced?) even the simplest cell to become unimaginably complicated. Just as two objects moving at random in a three-dimensional space need never meet, perhaps molecular biology's leap from one to three dimensions will take it beyond human comprehension. The old-timers must often look back longingly to that sunlit June evening, when molecular biology was a self-confident teenager and Crick a stripling youth of 50. □

John Cairns is at Hollygrove House, Wilcote, Charlbury OX7 3EA, UK.
e-mail: j.cairns@cts.u.ox.ac.uk

Perhaps molecular biology's leap from one to three dimensions will take it beyond human comprehension.



Hotchkiss reads his poem, "A Happy Crickmass", as Crick (far left) and the other party-goers look on.

Electrons seen in orbit

Colin J. Humphreys

The classic textbook shape of electron orbitals has now been directly observed. As well as confirming the established theory, this work may be a first step to understanding high-temperature superconductivity.

The world-famous bongo drums player, Richard Feynman, who also won the Nobel Prize for Physics, once asked the following question: if you had only one sentence in which to pass on to the next generation the most important scientific knowledge we possess, what would that sentence be? Feynman's own answer was "everything is made of atoms". "But what is their size and shape," the next generation might reply, "and how do they stick together?"

"Well," Feynman might have said "most scientists think of an atom as a sphere, like a tiny orange, and if you magnify an orange to the size of our Earth then every atom in the orange is magnified to the size of the original

orange. But in reality, many atoms are not spherical and the shape of atoms critically affects how they bond together."

On page 49 of this issue, Zuo *et al.*¹ use a combination of electron and X-ray diffraction to study the shape and bonding of oranges — in this case, copper atoms in copper oxide. For the first time the striking shape of some of the electron orbitals is revealed experimentally. These methods could be used to determine bonding in high-temperature copper oxide superconductors, the theory of which has eluded some of the world's best scientists for more than ten years.

What do we mean by the size and shape of an atom? In the simplest atom, hydrogen, a

single electron at a distance r from the nucleus is attracted to the nuclear charge e^+ by a potential $V = -e^2/r$. Solving the quantum-mechanical Schrödinger equation for this potential tells us which electron states are allowed. Because V has spherical symmetry about the nucleus, the simplest solutions also have spherical symmetry. The lowest energy state, or ground state, is called the $1s$ orbital. So, the hydrogen atom in its ground state may be visualized as being a spherical charge distribution centred on the nucleus (Fig. 1a)².

There are also higher energy states, both spherically symmetrical and less symmetrical. The simplest states with lower symmetry are the $2p$ states, $2p_x$, $2p_y$ and $2p_z$, which all have the same energy and so we say they are degenerate (Fig. 1b). But if the three states are added together they make a spherically symmetric distribution. The next simplest degenerate states are the $3d$ states, of which there are five (Fig. 2). The $3d$ states are particularly important in transition elements, such as copper, because they are only partially filled, and hence a copper ion can exhibit a variety of charges. Whereas in an isolated hydrogen atom all the $3d$ states have the same energy, this degeneracy is removed in transition elements if they are in an environment which is not symmetrical, as is the case in many crystals. The shape and position of the 'lobes' of the $3d$ orbitals are crucial for understanding the directionality of bonding in transition-metal compounds.

The paper by Zuo *et al.*¹ is remarkable because the quality of their charge-density maps allows, for the first time, a direct experimental 'picture' to be taken of the complex shape of the d_{z^2} orbital (Fig. 2). The correspondence between the experimental charge-density map of Zuo *et al.* (see Fig. 3c on page 50) and the textbook d_{z^2} orbital is striking. In the past, charge-density maps from crystals have usually been measured using X-ray diffraction, which accurately reveals the main peaks of the charge density (that is, the atom positions). For example, Crick and Watson famously interpreted the X-ray diffraction results of Rosalind Franklin to obtain the double-helix structure of DNA.

Although X-ray diffraction reveals peak positions it is normally unable to give details about the shape of the charge distribution, in particular the shape of the bonds. The main reason for this is that crystals contain defects

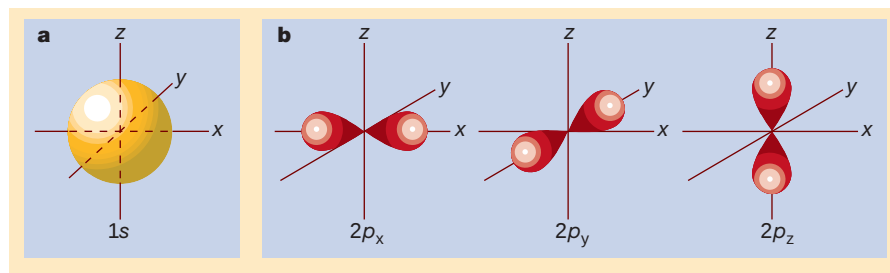


Figure 1 Textbook examples of electron orbitals. a, The ground state of the hydrogen atom. The $1s$ orbital is represented by a sphere centred on the nucleus. b, The higher energy $2p$ orbitals. The $2p_x$, $2p_y$ and $2p_z$ orbitals all have the same energy, and the orbital direction is usually dictated by the environment, such as the position of neighbouring atoms in a crystal.

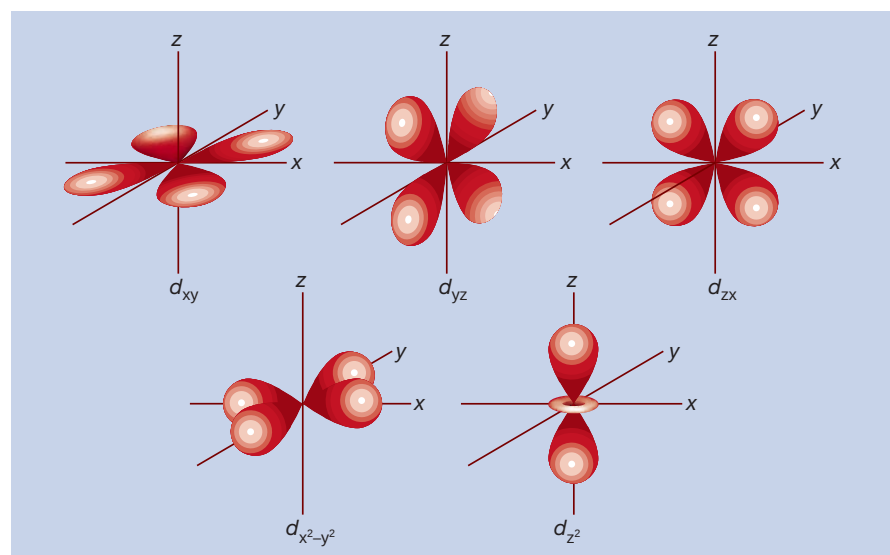


Figure 2 The family of $3d$ orbitals. Despite appearances, the five orbitals are all equivalent, and when superimposed add up to a spherical distribution. Direct observation of the d_{z^2} orbital in a copper oxide by Zuo *et al.*¹ may help answer questions about bonding in the high-temperature copper oxide superconductors.

such as dislocations, and the X-ray scattering from such defects is greater than the scattering from bonding electrons. But using an electron microscope it is possible to image the crystal, select a perfect region between crystal defects, and form a diffraction pattern from this perfect region. Zuo *et al.*¹ use an electron beam focused to nanometre dimensions to find a region of the crystal where the perfect-crystal theory of diffraction applies. They also filtered the electrons transmitted by the specimen to remove inelastically scattered electrons and ease comparison with theory. Using these methods, very accurate charge-density maps can be made which reveal the shape of the electron bonds.

In metal oxides, the metal ions interact mainly with the oxygen atoms. But in many copper- and silver-oxide compounds the metal ions are found in close-packed crystal structures, leading to predictions of short-range metal-metal bonding. Such bonding would be covalent, and therefore require non-spherical orbitals, unlike the usual ionic bonding, which involves electrostatic interactions between spherical ions. The non-spherical charge density observed by Zuo *et al.* around Cu atoms in Cu₂O reveals unusual *d*-orbital holes (where the *d_{z²}* orbital is unoccupied) and provides strong evidence for Cu-Cu bonding.

How does this work relate to high-temperature superconductivity? We now know that in the simple oxide Cu₂O there are *d*-orbital holes located on the Cu atoms. We also know that in the copper-oxide super-

conductors, which have CuO₂ (rather than Cu₂O) conducting planes, the charge carriers in the normal state are holes and in the superconducting state are hole pairs. In the CuO₂ superconductors, the available evidence suggests that the holes are on oxygen sites rather than on copper sites^{3,4}, whereas the work of Zuo *et al.* shows that for Cu₂O the holes are on copper sites. It would be of great interest to extend this approach to measuring charge densities to the more complex CuO₂ superconductors. Such an experiment should give us the following information about high-temperature superconductors: first, whether the holes are on copper or oxygen sites; second, how many holes there are per CuO₂ unit in the conduction planes and how many holes are elsewhere in the structure. Finally, we would like to know whether the distribution of holes in the conduction planes is actually periodic, as is usually assumed, or if there is an irregular two-phase charge density of holes as recently proposed⁵. Zuo *et al.*¹ have provided us with a tool to answer all these questions. ■

Colin J. Humphreys is in the Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK.

e-mail: cjh1001@hermes.cam.ac.uk

1. Zuo, J. M., Kim, M., O'Keeffe, M. & Spence, C. J. *Nature* **401**, 49–52 (1999).
2. Cottrell, A. *Introduction to the Modern Theory of Metals* (Institute of Metals, London, 1988).
3. Temmerman, W. M. *et al.* *J. Phys. C* **21**, L867–L874 (1988).
4. Nücker, N. *et al.* *Phys. Rev. B* **53**, 14761–14764 (1996).
5. Humphreys, C. J. *Acta Crystallogr. A* **55**, 228–233 (1999).

Ecology

Bite the mother, fight the daughter

Erkki Haukioja

Induced defences — which emerge or develop to full strength when activated by the presence of an enemy — are widespread in sedentary or slowly moving prey¹. Such induced resistance, as opposed to constitutive resistance (which is present all the time), is profitable when defences are metabolically expensive, and when attack is unpredictable but recurrent². Induced defensive traits are diverse. Water fleas, for example, form long, helmet-shaped spines on their necks, making it difficult for predators to handle them. Plants, on the other hand, can produce high levels of toxic compounds in their leaves. Much of the literature on induced plant defences contains reports of predators having a bad time of it if they try to target a plant that has already been attacked. However, these observations do not exclude the possibility that the poor quality of previously damaged plants is a fortuitous by-product of recovery from

attack, not a sign of increased defence.

This is not the case for the organisms described by Agrawal *et al.*³ on page 60 of this issue. These authors have used radish plants (*Raphanus raphanistrum*) and water fleas (*Daphnia cucullata*) to provide the best evidence yet that predator-activated responses



Figure 1 Radish plant under attack by the butterfly larva *Pieris rapae*. The plant's defence response includes the production of bitter-tasting, toxic compounds, and the development of hairy trichomes.

help the victims to survive an attack⁴. Not only that, but they show that these defences can be transmitted from an organism to its offspring. Induced defence is known to be a form of genotypic plasticity — that is, the ability of a given gene sequence (genotype) to produce various traits (phenotypes) depending on environmental conditions. But it has never previously been reported to extend across generations.

To be ready when needed, induced traits must develop quickly in the presence of a predator. This might be a real problem. Fortunately, it can be overcome if the target organism receives cues of the predator before the actual attack, or if the first attack is unlikely to be lethal. Although plant defences may be induced simply because herbivores are chewing the foliage, many organisms have sophisticated early-warning systems. Some plants, for example, switch on their defence mechanisms as soon as they detect volatile chemicals released by damaged plants close by⁵. Some of these volatiles also directly lure predators to attack plant-eating insects or mites⁶. But the first encounter with a predator is likely to be fatal for young water fleas, so they develop long neck spines in response to chemical cues emitted from their invertebrate predators. In the field, this causes water fleas to develop spiny, defensive helmets when living in a pond also inhabited by predators.

Agrawal and co-workers³ now report that in both wild radish, which is a cruciferous plant (its petals are arranged in a cross shape), and the water flea, the risks of a young individual's first encounter with predators can be reduced by the experiences of their parents. In the case of the wild radish, the authors allowed butterfly larvae (*Pieris rapae*; Fig. 1) to attack the leaves in a controlled manner. The radish plants responded by a ten-fold increase in the concentration of glucosinolate, and they also developed more numerous hairy trichomes on their leaves. Glucosinolates are bitter-tasting, toxic compounds that reduce insect feeding; hairy trichomes on leaves have a similar effect. The authors then exposed seedlings produced by these plants to further larval predation. They found that the offspring of damaged mothers provided less suitable diets for larvae than seedlings from undamaged, control plants.

A similar thing happened when Agrawal *et al.* exposed water fleas (Fig. 2) to chemical cues called kairomones from two invertebrate predators — the water fleas developed long helmets. Furthermore, the offspring of kairomone-treated mothers produced longer helmets than offspring from control mothers, irrespective of the environment in which the offspring were raised. The authors saw the same effect in successive broods produced later by the kairomone-treated mothers in clean water. This result confirms that

the development of long helmets in the embryonic stages resulted from maternal effects, and did not need to be directly induced by kairomones. However, offspring produced by the kairomone-treated mothers in clean water had shorter helmets than offspring raised in water that contained the chemical. This indicates that a recurrent induction is needed for a maximum-strength defensive response.

At first glance, the observation that certain traits are carried over to the next generation looks like an example of Lamarckian inheritance. This theory, proposed by the nineteenth-century botanist and zoologist Jean-Baptiste Pierre Antoine de Lamarck, holds that organisms can pass on to their offspring traits that they have acquired during their lifetime. Lamarck's theory is surely wrong if it is interpreted as meaning, say, that mice with cut tails would have offspring with no tails, or that people with trained muscles would produce muscly babies. But if the parent's environment affects how the genetic code in the offspring is translated, then certain acquired traits can be delivered to the offspring.

Accordingly, the most parsimonious explanation of Agrawal and colleagues' findings is that the genes that were switched on in the parent to generate the defensive response are also switched on in the offspring. The internal milieu of the radish seeds was indeed different between induced and control mothers, the induced mothers containing higher concentrations of glucosinolate. But this does not need to be the decisive

switch — several alternative mechanisms, passed on from either the mother or the father, can lead to the inheritance of environmentally affected traits⁷. Because the genetic code inscribed in the chromosomes does not change, however, the inter-generational carry-over effect is presumably reversible.

Nonetheless, the demonstration that chemically mediated defensive traits can be carried over between generations in two totally unrelated species hints that such phenomena might be widespread. Interest is growing in the idea that certain environmental effects can be inherited in many types of organism, from bacteria to plants, insects and mammals⁷. Although it is not yet clear to what extent these inter-generational responses are adaptive (that is, tailored to equip an organism for a particular situation), they have been reported in cases when parents and offspring meet similar environmental challenges. For example, parental crowding alters the physiology, behaviour, coloration and structures of the offspring in some species of locust⁸ and lepidopteran⁹. This helps them to manage at high or low

population densities — situations that often occur over several successive generations. Such observations indicate that if we do not take into account the environmental fates of the parents of experimental organisms, we may find unexplained variance in the results of ecological experiments. This is particularly true for research on induced defences, and Agrawal and colleagues' study indicates that ecologists have much to do to uncover sources of variation, such as how dynamic or long-lasting the effects of induced defences can be. ■

Erkki Haukioja is in the Department of Biology, University of Turku, FIN-20014 Turku, Finland.
e-mail: haukioja@utu.fi

1. Harvell, C. D. *Q. Rev. Biol.* **65**, 323–340 (1990).
2. Baldwin, I. T. *Proc. Natl Acad. Sci. USA* **95**, 8113–8118 (1998).
3. Agrawal, A. A., LaForsch, C. & Tollrian, R. *Nature* **401**, 60–63 (1999).
4. Agrawal, A. A. *Science* **279**, 1201–1202 (1998).
5. Karban, R. & Baldwin, I. T. *Induced Responses to Herbivory* (Univ. Chicago Press, 1997).
6. Dicke, M. et al. *J. Chem. Ecol.* **19**, 581–599 (1993).
7. Rossiter, M. C. *Annu. Rev. Ecol. Syst.* **27**, 451–476 (1996).
8. Applebaum, S. W. & Heifetz, Y. *Annu. Rev. Entomol.* **44**, 317–341 (1999).
9. Rose, D. J. W., Dewhurst, C. F. & Page, W. W. *Integr. Pest Manage. Rev.* **1**, 49–64 (1995).

Random fluctuations

Unsolved problems of noise

Peter V. E. McClintock

Random noise, and its effect on physical systems, has always been an interdisciplinary subject. Using the study of fluctuations as a unifying theme can lead to interesting scientific conferences, such as the recent UPoN'99 meeting in Adelaide, Australia*.

Over the years, noise has attracted some of the best scientists: in 1828 the botanist Robert Brown¹ discovered the random fluctuations of tiny particles in a fluid, which remained a mystery until Einstein's epoch-making paper² on Brownian motion in 1905. But it was not until later that the ubiquitous presence of noise began to be appreciated. A milestone was Johnson's 1927 paper³ in which he identified noise in electrical circuits as being of thermal origin. In the middle of the twentieth century, noise seemed merely a nuisance — to be eliminated or minimized wherever possible. It spoiled radio reception and musical recordings, and limited the precision of physical measurements. Later, it was found to wash out the fine structure of beautiful fractal patterns generated by nonlinear systems.

But noise is much more than just a nuisance. By the 1970s it became apparent that noise can play a creative role too. One of the

first examples was the observation of noise-induced transitions⁴, whereby the state of a nonlinear system can be utterly changed by the introduction of noise above a critical intensity. Another was stochastic resonance, in which a weak periodic signal in a nonlinear system can be amplified by added noise. Stochastic resonance was introduced^{5,6} to try to explain the Earth's ice-age cycle, but it is now recognized to be far more common, occurring for example in lasers, electronic circuits and sensory neurons. Yet another creative effect of noise is seen in Brownian ratchets⁷ where a net current of particles can be driven by noise, providing that there is an appropriate asymmetry in the system: this particular phenomenon may underlie the transport of macromolecules within biological cells. And, of course, it is noise in the sense of thermal fluctuations that drives chemical reactions, including those responsible for life itself.

The negative aspects of noise are still with us and, given its Janus-like character, how is noise to be perceived, pursued and investigated? A proper appreciation of noise must involve studies of both the mechanisms that produce it, for example in semiconductor devices, and the effect that it has on systems subject to it. Both these aspects were well represented at UPoN'99.

Of the several kinds of Brownian ratchet



Figure 2 Mother knows best. Agrawal *et al.*³ have shown that defence responses induced in response to predators in female water fleas (pictured carrying eggs) can be passed on to their offspring.

*Second International Conference on Unsolved Problems of Noise and Fluctuations (UPoN'99), Adelaide, Australia, 11–15 July 1999. Proceedings to be published by the American Institute of Physics.

that have been considered, the 'flashing ratchet' has attracted particular attention because of its likely relevance to biology⁷. In this ratchet, a noise-driven particle — which might be attached to a vesicle carrying a macromolecule — diffuses within a periodic potential (a cyclic variation of the particle's potential energy with position). This potential can typically take one of two forms — it can either be flat, with no variation at all, or it can have a series of ratchet-shaped teeth. The system flashes back and forth between these two configurations. As a result — depending on the shape of the teeth and the characteristics of the noise — the particle will diffuse preferentially either to the right or to the left; given noise of appropriate intensity, there will be a net current. Remarkably, there is still a net current of particles, albeit a smaller one, even if the potential is tilted so that the particles have to run uphill. Yet if the flashing ceases, the system stops working. Current will not flow uphill for either of the two configurations taken alone.

Surprising though it may seem, there appears to be a close relationship between flashing ratchets and certain mathematical games. Juan Parrondo (Universidad Complutense de Madrid, Spain) has devised two simple games whose behaviour was shown by Derek Abbott (University of Adelaide and organizer of *UPoN'99*) and colleagues to model closely that of a flashing ratchet. Either game played on its own results in loss for the player. But if play alternates randomly between the two games the result is a win — that is, two losing strategies can result in a win. The games are amenable to rigorous mathematical analysis, and promise to pave the way to a detailed understanding of ratchet phenomena. Such models could have a bearing on population genetics and evolutionary biology, as well as on game theory in economics.

In surveying stochastic resonance, Sergey Bezikov (National Institutes of Health, Bethesda, Maryland) admitted ruefully that early perceptions now seem inadequate. In particular, the once standard definition⁸ of stochastic resonance as an effect confined to bistable systems that can exist in two different states (including systems with thresholds) is not very useful because the phenomenon occurs much more widely. In fact, the only general definition of stochastic resonance that remains tenable is that the effect arises whenever a nonlinear system has a large, strongly noise-dependent susceptibility⁹ (a quantity that relates the periodic response of the system to the weak periodic force that gives rise to it).

What do stochastic resonance, Brownian ratchets and, indeed, most other fluctuational phenomena have in common? It has slowly been realized that a unifying feature in many cases is reliance on 'special' fluctuations that occur infrequently. For example, a

system will usually wait for a long time below a potential barrier for a fluctuation of exactly the right shape and strength to carry it over. These large, rare fluctuations are of fundamental importance in understanding the behaviour of noise-driven systems. They are starting to be understood, even in the dauntingly complex, non-equilibrium cases found in practical situations.

Advances have recently been made in experimenting on large, rare fluctuations. Riccardo Mannella (University of Pisa, Italy) described experiments that have confirmed the validity of the so-called logarithmic susceptibility, a new physical quantity¹⁰ for describing large fluctuations in periodically driven non-equilibrium systems. These results are immensely encouraging because, if large fluctuations can be properly understood, it may then be possible to control their destructive power (for example, in the spon-

taneous failure of lasers) and, perhaps, to exploit them (such as when minimizing the energy requirements for desired transitions). Noise is starting to look altogether friendlier and more useful than it did in earlier times. ■

Peter V. E. McClintock is in the Department of Physics, University of Lancaster, Lancaster LA1 4YB, UK.

e-mail: p.v.e.mcclintock@lancaster.ac.uk

1. Brown, R. *Phil. Mag.* **4**, 161 (1828).
2. Einstein, A. *Ann. Phys.* **17**, 549 (1905).
3. Johnson, J. B. *Nature* **119**, 50–51 (1927).
4. Horsthemke, W. & Lefever, R. *Noise-Induced Transitions* (Springer, Berlin, 1984).
5. Benzi, R., Sutera, S. & Vulpiani, A. *J. Phys. A* **14**, L453–L457 (1981).
6. Nocolis, C. *Tellus* **34**, 1–9 (1982).
7. Bier, M. *Contemp. Phys.* **38**, 371–379 (1997).
8. Gammaitoni, L., Hänggi, P., Jung, P. H. & Marchesoni, F. *Rev. Mod. Phys.* **70**, 223–288 (1998).
9. Dykman, M. I., Mannella, R., McClintock, P. V. E. & Stocks, N. G. *Phys. Rev. Lett.* **65**, 2606 (1990).
10. Smelyanskiy, V. N., Dykman, M. I., Rabitz, H. & Vugmeister, B. E. *Phys. Rev. Lett.* **79**, 3113–3116 (1997).

Neurobiology

Young receptors make smart mice

T. V. P. Bliss

Animals can learn because of changes in the brain which allow new information to be acquired, stored and later recalled. But what are these changes and where do they occur? At the cellular level, changes probably occur at synapses — the junctions between nerve cells. This idea is formalized in the Hebb rule, which states that a synapse between cell A and cell B will be strengthened if the two cells are active at the same time¹. Neuroscientists can explore the properties and behavioural implications of the Hebb synapse by studying an experimental model of synaptic plasticity known as long-term potentiation (LTP)². Synapses that show LTP

are found in several parts of the brain, notably in the hippocampus (a cortical structure which, in humans, is required for the formation of autobiographical memories). Does LTP, then, supply the synaptic underpinning of memory? A report by Tang *et al.*³ on page 63 of this issue provides new evidence in support of this hypothesis, showing that LTP is considerably enhanced in transgenic mice with improved learning performance.

Long-term potentiation is induced by brief, repeated stimulation of defined neural pathways in the hippocampus. The resulting increased efficacy of synaptic transmission

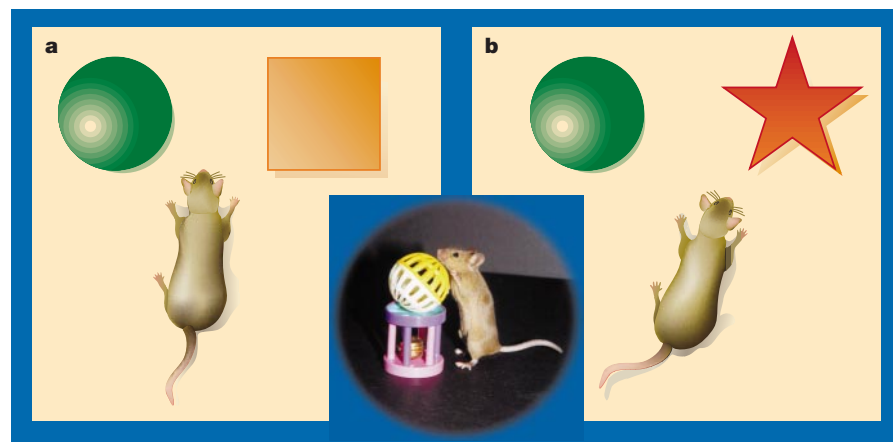


Figure 1 Object-recognition task. **a**, In an initial training session the mouse explores two objects in a box, devoting roughly equal time to each. **b**, When the mouse is then re-exposed to one of these objects, together with a new object, it spends more time exploring the new object. Tang *et al.*³ find that this bias is enhanced in transgenic mice (inset) overexpressing the NR2B subunit of the NMDA receptor, indicating improved recognition memory.

can last for hours (in slices cut from the hippocampus) or several days (*in vivo*). If drugs are used to block the induction of LTP, rats have trouble finding their way around a maze, suggesting that LTP is necessary for spatial learning⁴. And using transgenic mice, researchers can explore what happens to LTP and learning when specific proteins are overexpressed or deleted. When genes involved in regulating synaptic transmission are deleted, most of the mice can no longer sustain normal LTP. This impairment is usually accompanied by a defect in spatial learning.

To appreciate what Tang *et al.*³ have done, it is necessary to understand certain properties of a membrane protein called the NMDA (*N*-methyl-D-aspartate) receptor. This receptor binds to the excitatory neurotransmitter glutamate and controls the initiation of LTP in most hippocampal pathways. The receptor forms a channel that does not open unless two conditions are satisfied simultaneously — glutamate must bind to the receptor, and the membrane in which the receptor is embedded must be strongly depolarized. Once the channel opens a flux of calcium ions enters the cell, triggering the induction of LTP.

The NMDA receptor is made up of an NR1 subunit (Box 1), which is obligatory for channel function, and a selection of developmentally and regionally regulated NR2 subunits (A to D). The functional properties of the receptor depend on its subunit composition. For example, the glutamate-evoked current (which is important in determining the amount of Ca^{2+} that enters the cell) has a longer duration in receptors containing NR2B subunits than in those containing

NR2A subunits. The proportion of NR2B subunits is higher in young animals than in adults⁵, and this may account for the greater degree of LTP seen in young animals⁶. With this in mind, Tang *et al.* engineered a transgenic mouse that overexpresses the NR2B subunit in the adult brain. They then asked three questions. First, would the mutant show an increase in NMDA-receptor-mediated current? Second, would it, as an adult, show the enhanced level of LTP normally seen in young animals? And finally, would this make a smarter mouse? The answer to all three questions seems to be yes.

The authors successfully expressed the transgene in two different mouse lines, with similar physiological and behavioural effects in both. Glutamate-evoked NMDA-receptor-mediated currents were larger and of a longer duration in cultured neurons from transgenic than from normal animals. (Presumably, although direct evidence is not offered, this was also the case in the hippocampal slices in which LTP was monitored.) Tang *et al.* also found that LTP was greatly enhanced in the two transgenic lines, with a flat initial time course reminiscent of LTP in young rather than adult animals.

The consequences of overexpressing the NR2B subunit on learning and memory — or, at least, on those kinds of learning that depend on the hippocampus in rodents — seem unequivocal. The transgenic mice showed improved learning scores compared with normal mice in three different tests of their ability to acquire and retain information. In one of these tests, fear conditioning, a tone is paired with a mild foot-shock in a particular box. The mouse will freeze, indi-

cating memory of the shock, when it is replaced in the box some time later. This 'context specificity' is thought to depend on the hippocampus.

The authors found that associative learning was enhanced in the transgenic mice. Furthermore, when the animals were repeatedly given the tone alone or repeatedly returned to the box in which they had been trained, without being given further shocks, the conditioned response was extinguished more rapidly in the transgenic animals than in normal mice. Tang *et al.* regard this as further evidence for improved learning — an interpretation that requires extinction of the response to be seen as active relearning, rather than a passive forgetting of the previously learned association. In two other tasks (object recognition, shown in Fig. 1, and the water maze, in which animals have to locate a submerged platform in a pool of opaque water), the transgenic animals again scored more highly than normal mice.

Although this is not the first report of a transgenic animal showing both enhanced LTP and improved learning⁷, Tang and colleagues' work is striking for two reasons. First, there is the clear link between manipulation of a specific gene (overexpression of the NR2B subunit) and the physiological consequence (enhanced Ca^{2+} flux, leading to greater LTP). The second reason is the variety of tasks examined, which adds strength to the authors' conclusion that they have indeed made a smarter animal.

An improvement in LTP has been reported in other transgenic mice. When the genes encoding postsynaptic density-95 (PSD-95; a cytoplasmic protein that binds the NMDA receptor)⁸ and GluR2 (a subunit of the AMPA subtype of glutamate receptor)⁹ were deleted, both showed enhanced LTP. Yet their performance in the water maze was impaired. Compare this with Tang and colleagues' mice, which showed an improved performance in the water maze. The PSD-95 knockout mice were also defective in another form of activity-dependent synaptic plasticity, long-term depression (LTD). By contrast, LTD appeared normal in Tang and colleagues' mice. This may explain the different responses in the water maze — bidirectional changes in synaptic efficacy are thought to be important for efficient storage in neural nets¹⁰.

Several genetically targeted mice have been described in which a reduction in LTP has not been matched by diminished performance in hippocampal-dependent learning (see, for example, ref. 11). Against this background, where changes in LTP correlate with changes in learning in some transgenic animals but not in others, it is not surprising that a consensus on the role of synaptic plasticity in learning has been hard to achieve¹². But the impressive data obtained by Tang *et al.* will bolster the majority view that synap-

Box 1: The NMDA receptor and schizophrenia

In the work discussed in the main article here, Tang *et al.*³ show that mice in which the NR2B subunit of the NMDA receptor is overexpressed display a striking increase in synaptic plasticity and learning ability. But what about mutations of the essential NR1 subunit? Mice with targeted ablation of this subunit lack functional NMDA receptors and are not viable. However, newly developed mutants, genetically engineered to express about 5% of the normal number of NR1 subunits, live to maturity. And, according to a report by Mohn *et al.*¹³ in the latest issue of *Cell*, these mice may provide the best rodent model of schizophrenia so far.

Mohn and colleagues find that reduced expression of the NR1 subunit (and, hence, the NMDA receptor) leads to increased motor activity and rapid repetitive behaviour (stereotypy), as well as reduced social and sexual interactions. These 'symptoms' are all characteristics of pharmacologically induced models of schizophrenia.

The finding that a reduced number of NMDA receptors leads to schizophrenic types of behaviour should help to distinguish between two rival theories as to the causes of the disease. The first, the 'dopamine hypothesis', derives from the observation that antipsychotic drugs share a common action in blocking

receptors for the neurotransmitter dopamine. This theory attributes the behavioural and cognitive abnormalities of schizophrenia to increased levels of dopamine. The second theory, the 'glutamate-dysfunction hypothesis', evolved from the observation that phencyclidine, an antagonist of the NMDA receptor, produces symptoms that mimic schizophrenia in people and mice. According to this hypothesis, the disease is caused by a disturbance in the balance between the reciprocal actions of glutamate and dopamine on subcortical structures¹⁴. So the results of Mohn *et al.* support the glutamate-dysfunction hypothesis.

T. V. P. B.

tic plasticity is an essential component of the neural processes underlying learning and memory in the vertebrate brain. ■

T. V. P. Bliss is in the Division of Neurophysiology, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK.

e-mail: tbliss@nimr.mrc.ac.uk

1. Hebb, D. O. *The Organization of Behavior* (Wiley, New York, 1949).
2. Bliss, T. V. P. & Collingridge, G. L. *Nature* **361**, 31–39 (1993).
3. Tang, Y.-P. et al. *Nature* **401**, 63–69 (1999).
4. Morris, R. G. M., Anderson, E., Lynch, G. S. & Baudry, M.

Nature **319**, 774–776 (1986).

5. Monyer, H., Burnashev, N., Laurie, D. J., Sakmann, B. & Seeburg, P. H. *Neuron* **12**, 529–540 (1994).
6. Harris, K. M. & Teyler, T. J. *J. Physiol. (Lond.)* **346**, 27 (1984).
7. Manabe, T. et al. *Nature* **394**, 577–581 (1998).
8. Migaud, M. et al. *Nature* **396**, 433–439 (1998).
9. Jia, Z. P. et al. *Neuron* **17**, 945–956 (1996).
10. Willshaw, D. & Dayan, P. *Neural Comp.* **2**, 85–93 (1990).
11. Zamanillo, D. et al. *Science* **284**, 1805–1811 (1999).
12. Martin, S. J., Grimwood, P. D. & Morris, R. G. M. *Annu. Rev. Neurosci.* (in the press).
13. Mohn, A. et al. *Cell* **98**, 427–436 (1999).
14. Carlsson, M., Hansson, L. O., Waters, N. & Carlsson, M. L. *Life Sci.* **61**, 75–94 (1997).

Dark matter

Lumpy haloes spin faster

James Binney

More than 90% of the mass of the Universe is thought to be made up of dark matter. The density of this mysterious matter has been most precisely measured in the outer reaches of disk-shaped spiral galaxies, where very little light is emitted by stars. It is generally accepted that galaxies are surrounded by dark matter ‘haloes’ in which the visible galaxy is embedded, so helping to explain the rapid rotation of stars and gas far from the galactic centre. How much dark matter coexists with stars near the centre of galaxies is less certain, but numerical studies of the dynamics of bar-shaped galaxies have been used to impose an upper limit on the central dark-matter density¹.

Tremaine and Ostriker² now suggest in *Monthly Notices of the Royal Astronomical Society* that, because of an unexpectedly strong interaction with the galactic disk, dark haloes may have a rapidly spinning central bulge, similar to the stellar cores of galaxies. If this is true, it will come as a relief to advocates of the ‘cold dark matter’ (CDM) picture of the Universe, which predicts that many disk galaxies should be dominated by dark matter even at their centres.

For galaxies to shine, some of their mass must be luminous baryons, rather than dark matter. According to the CDM model, during galaxy formation this baryonic matter loses energy through radiation and therefore collapses into a compact, rapidly spinning disk near the centre of a much larger and barely rotating dark-matter halo. Rapidly spinning, self-gravitating bodies are prone to becoming rotating bars, and most disk galaxies have bars at their centres. More than ten years ago, Weinberg³ pointed out that a rapidly spinning bar embedded in a slowly rotating dark halo would transfer angular momentum to the halo through dynamical friction. In this way the rotation of the halo would be accelerated, but the bar would be slowed down. So, if the density of dark matter is high at the centre of galaxies, bars should rotate slowly. But observations show

that most bars are still rapidly rotating, suggesting that dark matter is unimportant in the centres of galaxies. Although Weinberg’s arguments were mostly analytical, they were later confirmed by numerical simulations¹.

Tremaine and Ostriker² weaken this line of reasoning by arguing that a disk as well as a bar may be able to cause the halo to spin faster, in which case the bar will not be severely slowed by the halo. Non-axisymmetric features of the disk, such as the spiral structure, are not massive enough to transfer significant angular momentum to the halo. But Tremaine and Ostriker suggest that the disk and halo may be effectively coupled by inhomogeneities in the halo, which gravitationally concentrate the orbits of stars and gas in the disk. This creates regions of higher disk density, which in turn gravitationally pull the halo irregularities around in the sense of the disk’s rotation.

Are the interiors of dark haloes sufficiently lumpy for effective coupling of the disk and halo as proposed by Tremaine and Ostriker? Until recently the standard picture of a dark halo was a smooth object with a radial density gradient that decreased from the outside inwards. As numerical simulations of halo formation grew more detailed, it became apparent that, in the CDM picture at least, haloes must be full of high-density blobs and strings⁴. This is because structure formation is a ‘bottom-up’ process, starting with the formation of very low-mass haloes, many of which merge to form middle-size haloes, and so on. When small haloes merge to form a more massive halo, the outer layers of the haloes are quickly stripped off by tidal forces, whereas the cores stay intact for much longer. In addition, as a stripped core falls through a more massive halo, it loses a ribbon of particles — a ‘tidal streamer’ — that moves on essentially the same orbit as the parent halo. Tremaine and Ostriker estimate the number of tidal streamers and intact halo cores that are needed to ensure sufficient transfer of angular momentum between disk and halo.



100 YEARS AGO

On removing some virgin cork from the wall of a conservatory a short time ago, I was much struck with the way in which a small black female spider clung to her two egg-bags, despite the fact that the piece of cork to which she was clinging had been thrown roughly to the ground. ... Knowing how fully alive to danger the spider race is in general, I thought that this remarkable instance of devotion to maternal promptings on the part of a naturally sensitive creature ought not to be disregarded. I accordingly removed the mother very carefully, and placed her on some rockwork, where I noticed she seemed to be very uneasy, moving restlessly about as if searching for something. I then took the egg-bags and placed them beside her. As I expected, she seemingly failed to recognise them, or at least manifested a repugnance to them, and ran away for a little distance. Subsequently, however, she returned, and proceeded to examine the bags with scrupulous care by means of her palpi; and evidently satisfied with this scrutiny that they were really her own cherished property, she commenced to spin a web about them to secure them in their place.

From *Nature* 31 August 1899.

50 YEARS AGO

There is an old Egyptian saying that whoever has once drunk of the waters of the Nile will want to go back and drink them again. I do not know if that has ever been said of the waters of the Tyne, but it is certainly true that the British Association always welcomes an invitation to Newcastle. ... We came first in 1838 when England was mainly an agricultural country. The 111 years that have elapsed have seen profound changes in our national life, particularly the prodigious advancement of national science and of its daughter technology. Among the many resulting achievements has been a great increase in the certainty and the length of life. The world population has risen more rapidly than ever before; it is now estimated by the Food and Agriculture Organisation at about 2,300 millions, and the increase at about 20 millions a year — an average addition of two every three seconds, day and night, year after year, and the two may become three or more as science advances, social services improve, and bodies such as the World Health Organisation become fully operative. Can they all be fed?

From *Nature* 3 September 1949.

The existence within galactic haloes of orbiting halo cores and tidal streamers is a general prediction of theories in which dark matter starts out cold. So it is vital to have some observational tests of these predictions. Two such tests are the frequency of galactic warps, and the amount by which the velocities of stars at a point in the galactic disk scatter about the velocity of a perfectly circular orbit — the disk's 'velocity dispersion'. Galactic disks are often warped far from the galactic centre, and the most plausible explanation of this phenomenon⁵ requires the existence of halo substructure just outside the stellar disk. Other observations place an upper limit on the magnitude of halo substructure within the outer radius of the stellar disk because draining angular momentum from a stellar

disk inevitably increases the disk's velocity dispersion. For example, the Sun lies just inside the radius at which the Milky Way's disk becomes warped, and the apparent low-velocity dispersion of the solar neighbourhood may turn out to be incompatible with the level of disk-halo coupling advocated by Tremaine and Ostriker. ■

James Binney is in the Department of Physics, University of Oxford, Keble Road, Oxford OX1 3NP, UK.

e-mail: j.binney1@physics.ox.ac.uk

1. Debattista, V. P. & Sellwood, J. A. *Astrophys. J.* **493**, L5–L8 (1998).
2. Tremaine, S. D. & Ostriker, J. P. *Mon. Not. R. Astron. Soc.* **306**, 662–668 (1999).
3. Weinberg, M. D. *Mon. Not. R. Astron. Soc.* **213**, 451–471 (1985).
4. Moore, B. *et al.* <http://xxx.lanl.gov/astro-ph/9907411>
5. Jiang, L.-G. & Binney, J. J. *Mon. Not. R. Astron. Soc.* **303**, L7–L10 (1999).

Protein unfolding

Trapped in the act

Tania A. Baker

Protein unfolding is critical for dismantling protein complexes, transporting proteins between cellular compartments and preparing proteins for destruction. On page 90 of this issue, Weber-Ban *et al.*¹ show that a model protein carrying a peptide tag that targets it for degradation is globally unfolded by a member of the Clp/Hsp100 protein family called ClpA. Similar unfolding reactions are probably involved in the many biological processes promoted by members of this family, and may be part of the mechanism used by many architecturally similar protein-digesting machines.

Members of the Clp/Hsp100 family are cylindrical complexes composed of six or seven subunits. They are involved in processes ranging from DNA replication and protein degradation to reactivation of heat-aggregated proteins and inheritance of prion-like factors (reviewed in ref. 2). Studies of bacterial ClpA and ClpX, and yeast Hsp104, have led to the hypothesis that members of the Clp/Hsp100 family recognize specific proteins, then promote a conformational change in these proteins using the energy of ATP hydrolysis. These Clp/Hsp100-catalysed changes in protein conformation have many biological consequences. For example, several 'protein remodelling' reactions have been reported^{3,4}, in which multimeric protein complexes are disassembled into their component subunits to trigger a biological response.

Some Clp/Hsp100 proteins, including ClpA, also act directly in protein degradation. This process is critical for normal cellular function, removing not only damaged proteins but also regulatory factors control-

ling other, fundamental processes such as progression of the cell cycle. To promote energy-dependent protein degradation, ClpA forms a complex with another cylindrical protein comprising 14 subunits of the ClpP peptidase⁵. ClpP is an ingeniously designed protein. Its many active sites, which cleave peptide bonds, are buried on the inside of the cylindrical chamber. Here they are shielded from proteins not scheduled for degradation⁶. Thus, in the ClpAP complex, ClpA's job is to recognize a substrate protein and translocate it into the ClpP chamber for destruction⁷. Current structural information indicates that the pore at the end of the ClpP ring is not wide enough to admit a globular folded protein, so this translocation also probably requires ClpA to alter the conformation of its substrate.

How does ClpA choose proteins for destruction? One tagging system, which specifically marks incomplete proteins for degradation, has been characterized in bacteria⁸. Tagging occurs during protein synthesis. When the synthetic machinery stalls — because the messenger RNA being translated is broken and lacks a proper stop signal — a small RNA called SsrA is recruited to the site of synthesis on the ribosome. SsrA provides the necessary stop sign to release the stalled ribosome and, as part of this rescue process, it directs the addition of 11 amino acids to the end of the newly synthesized protein fragment. This 11-residue tag is a powerful protein-destruction signal that is recognized by several proteases, including the ClpAP complex.

Weber-Ban *et al.*¹ now show that ClpA can change the conformation of a substrate protein, unfolding it and then passing it

directly to ClpP (Fig. 1). The authors constructed an SsrA-tagged version of green fluorescent protein (GFP), which, in its native, folded state, fluoresces green. In this way, they could easily monitor unfolding of the GFP by a decrease in fluorescence, and the GFP's SsrA tag converts it into a substrate for degradation by ClpAP. When the authors incubated GFP–SsrA with ClpA (in the absence of ClpP), they observed a small, temporary — but detectable — drop in fluorescence. This result indicates that ClpA might unfold the protein, but that GFP rapidly refolds and regains its glowing, green state.

To overcome this refolding problem, the authors used a variant of another protein machine that influences pathways of protein folding. This machine, the molecular chaperone GroEL, specifically binds unfolded proteins by interacting with patches of hydrophobic amino acids that are usually buried in the core of a folded protein. Certain mutant versions of GroEL can bind unfolded proteins but cannot release them⁹, so they irreversibly bind — or trap — unfolded proteins.

When Weber-Ban *et al.* added a GroEL

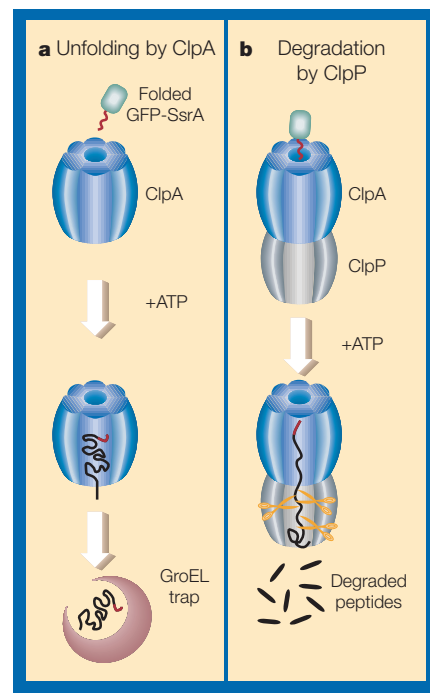


Figure 1 Protein unfolding and degradation by the ClpAP complex. ClpA is thought to unfold a target protein in preparation for destruction by the ClpP protease. Weber-Ban *et al.*¹ used green fluorescent protein (GFP) tagged with a destruction signal (a small peptide encoded by the SsrA RNA) as an unfolding substrate for ClpA. **a**, In the presence of ATP, ClpA unfolds the GFP–SsrA and the GFP stops fluorescing. The unfolded protein is irreversibly bound by a mutant form of the molecular chaperone GroEL, preventing it from refolding. **b**, In the presence of the ClpP, ClpA unfolds the GFP–SsrA. The target is then degraded inside the ClpP cylinder.

trap to ClpA and GFP-SsrA, the trap successfully captured unfolded GFP-SsrA generated by the action of ClpA. This result shows that ClpA alters the conformation of this substrate, converting it to a non-native state. Although binding to GroEL and loss of fluorescence both indicate that GFP-SsrA is significantly unfolded by ClpA, the authors confirmed this conclusion with an additional experiment. Using deuterium-hydrogen (^2H - ^1H) exchange, they showed that ClpA causes exposure of the entire polypeptide chain in GFP-SsrA to water, including regions of the GFP that are normally buried (and, hence, protected from solvent) by the three-dimensional state of the folded protein. Thus, ClpA has the capacity to globally unfold substrate proteins.

Global unfolding is likely to be a general feature during translocation of a substrate protein from ClpA into the proteolytic chamber of ClpP. Protein unfolding could also account for the roles of the Clp/Hsp100 proteins in other biological processes. However, it remains to be seen whether proteins that are remodelled or disassembled, rather than being degraded, are subject to such global unfolding. Moreover, ClpAP is a prototype for many proteolytic enzymes that have similar protein architectures, with

ATPase subunits flanking a cylindrical proteolytic chamber. The 26S proteasome, for example — the machine responsible for nearly all the protein degradation in eukaryotic cells — falls into this class. Perhaps the ATPases that make up the proteasome's 19S cap have a similar unfolding function to ClpA. The use of GroEL traps, and glowing substrates such as GFP, should facilitate the mechanistic exploration of these more complex protein machines. ■

Tania A. Baker is in the Department of Biology, Massachusetts Institute of Technology, and the Howard Hughes Medical Institute, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.

e-mail: tabaker@MIT.edu

1. Weber-Ban, E. U., Reid, B. G., Miranker, A. E. & Horwich, A. L. *Nature* **401**, 90–93 (1999).
2. Gottesman, A., Maurizi, M. R. & Wickner, S. *Cell* **91**, 435–438 (1997).
3. Schirmer, E. C., Glover, J. R., Singer, M. A. & Lindquist, S. *Trends Biochem. Sci.* **21**, 289–296 (1996).
4. Levchenko, I., Luo, L. & Baker, T. A. *Genes Dev.* **9**, 2399–2408 (1995).
5. Kessel, M. et al. *J. Mol. Biol.* **250**, 587–594 (1995).
6. Wang, J., Hartling, J. A. & Flanagan, J. M. *Cell* **91**, 447–456 (1997).
7. Hoskins, J. R., Pak, M., Maurizi, M. R. & Wickner, S. *Proc. Natl Acad. Sci. USA* **95**, 12135–12140 (1998).
8. Keiler, K. C., Waller, P. R. H. & Sauer, R. T. *Science* **271**, 990–993 (1996).
9. Fenton, W. A., Kashi, Y., Furtak, K. & Horwich, A. L. *Nature* **371**, 614–619 (1994).

Protein folding

Where do the electrons go?

Rudi Glockshuber

The biologically active, three-dimensional structure of a protein is its energetically most favourable conformation, and is encoded by its amino-acid sequence. Many proteins can fold correctly on their own (or with the help of molecular chaperones). But things are different for proteins containing disulphide bonds. Formation of a disulphide bond from the thiol (S-H) groups of two cysteine residues generates two protons and two electrons and changes the covalent structure of the polypeptide chain (Fig. 1). Consequently, these bonds cannot form spontaneously unless the protein interacts with an oxidant that accepts the electrons¹. Where do these electrons then go? Reporting in *Cell*, James Bardwell and colleagues² provide compelling experimental evidence that, in bacteria, oxidative protein folding is coupled to the electron-transport chain.

Why has it been so difficult to answer such an important question? The formation of disulphide bonds was originally thought to proceed directly from oxygen. But the pathway of electron flow that allows oxidative protein folding *in vivo* soon proved to be complex, involving a series of

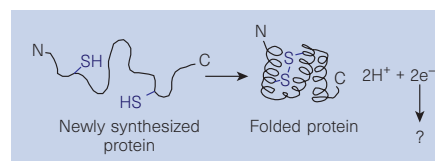


Figure 1 Formation of disulphide bonds during protein folding. Disulphide bonds form between sulphur atoms of the amino acid cysteine. They cannot form spontaneously unless the polypeptide chain interacts with an oxidant, which accepts the two electrons that are generated by the formation of each disulphide bond from two free thiol (S-H) groups.

specialized bacterial enzymes. Several years ago, Bardwell and others³ found that formation of disulphide bonds in the bacterial periplasm (the oxidizing compartment between the inner and outer bacterial membranes) requires an enzyme called DsbA. This enzyme randomly transfers its extremely reactive catalytic disulphide bond to folding proteins, and is itself recovered as an oxidant by disulphide exchange with a protein in the inner membrane called DsbB⁴ (Fig. 2). Then, last year, molecular oxygen was found⁵ to act as the final elec-

tron acceptor for DsbB-catalysed protein folding *in vitro*.

Can DsbB — which lacks cofactors such as haems or flavins — be oxidized directly by oxygen? The answer came from an intriguing observation. When a DsbB preparation was purified from 97% to more than 99% purity, the enzyme lost its ability to catalyse the oxidation of DsbA by molecular oxygen *in vitro*. But this pure DsbB then provided an assay for identifying the crucial components of the inner membrane that had been removed during the final purification step. Using this assay, Bardwell and co-workers² first found that the missing component is the haem-containing cytochrome *bd* oxidase complex — a hint that protein folding is linked to the electron-transport chain. Yet disulphide bonds still form normally in a cytochrome *bd* oxidase-deficient strain of *Escherichia coli*. So, the authors used membrane fractions of this strain to identify the cytochrome *bo* oxidase complex as an alternative component that can restore DsbB activity. These results mean that both terminal oxidases of the electron-transport chain in *E. coli* can independently transfer electrons from DsbB to molecular oxygen via their haem groups. When the authors analysed the membrane fractions from a strain deficient in both terminal oxidases they found that DsbB was not oxidized, showing that no other proteins can catalyse the flow of electrons from DsbB to oxygen.

Do both of these terminal oxidases interact directly with DsbB? Both are known to catalyse the flow of electrons from the ubiquinone pool in the inner membrane to molecular oxygen, suggesting that electrons might flow from DsbB to ubiquinone, then from ubiquinone, via the terminal oxidases, to oxygen⁶. Indeed, Bardwell and colleagues found that, *in vitro*, pure DsbB catalyses the oxidation of DsbA by ubiquinone derivatives extremely efficiently. The catalytic power of the purified terminal oxidases is also striking. In a completely reconstituted *in vitro* redox system, both were catalytically active at concentrations 1,000-fold lower than that of DsbA².

But this is still not the end of the story. Bardwell and co-workers² also found the pathway of electron flow that guarantees the formation of disulphide bonds in the absence of oxygen (Fig. 2). The existence of this pathway became evident when a strain of *E. coli* lacking both terminal oxidases and grown in the absence of oxygen showed no defects in disulphide-bond formation. It turns out that oxidized menaquinone — which can replace ubiquinone under anaerobic conditions and transports electrons to alternative electron acceptors such as fumarate or nitrate — can oxidize DsbB directly, albeit with a lower efficiency than ubiquinone². This menaquinone pathway also allows disulphide bonds to be formed

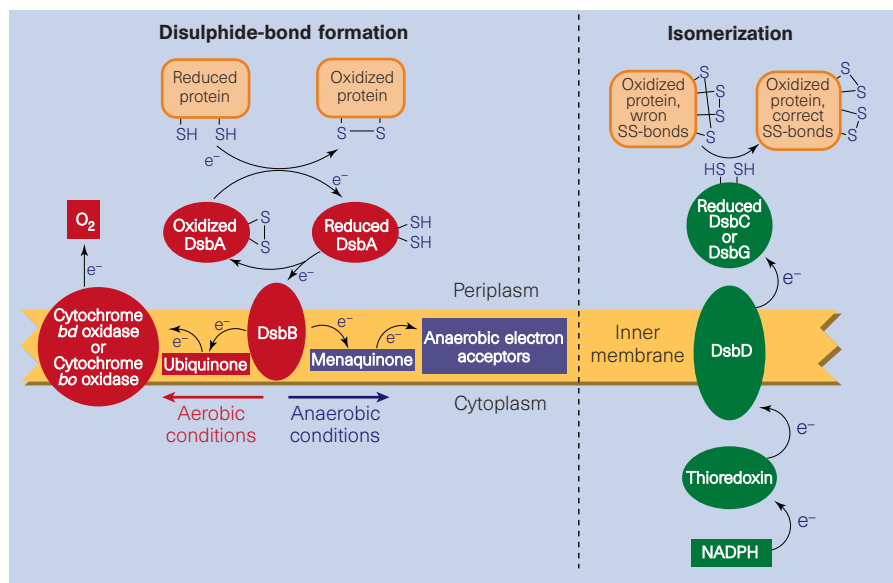


Figure 2 Electron-flow pathways for oxidative protein folding under aerobic and anaerobic conditions, as described by Bardwell and colleagues². An independent electron-flow pathway (right) catalyses isomerization of wrongly formed disulphide bonds in the periplasm. Electrons from NADPH in the cytosol keep the disulphide isomerases DsbC and DsbG in their catalytically active, reduced form through thioredoxin and DsbD. The reactions of DsbA, DsbC and DsbG with the folding polypeptide involve two electrons and occur via disulphide interchange reactions. The same applies to the oxidation of DsbA by DsbB, and the reduction of DsbC and DsbG by DsbD.

in strains that lack ubiquinone. But ubiquinone and menaquinone are probably the only oxidants in the *E. coli* inner membrane that can directly oxidize DsbB. Strains lacking ubiquinone and menaquinone show a strong defect in disulphide-bond formation⁶, and the same is true for an anaerobically grown strain in which fumarate cannot act as electron acceptor owing to the lack of fumarate reductase².

Bardwell and colleagues' study² answers many questions about how disulphide bonds form in bacteria. The pathway — in particular, the direct interaction between DsbB and quinones — is unexpected. Another surprise has been the recently discovered mechanism by which the DsbC or DsbG proteins catalyse isomerization of wrongly formed disulphide bonds in the periplasm^{7,8} (Fig. 2). These proteins must be kept reduced to be catalytically active as disulphide isomerases. This is accomplished by an inverse electron flow towards the periplasm, from NADPH in the cytosol, via thioredoxin to DsbD, and via DsbD across the inner membrane to DsbC and DsbG⁹.

That's all fine for bacteria, but how does oxidative protein folding occur in eukaryotic cells? Disulphide-bonded proteins generally cannot fold in reducing cellular compartments such as the cytoplasm. Instead, they fold in an oxidizing compartment, the endoplasmic reticulum. Oxidized glutathione has long been thought to act as the electron acceptor that reoxidizes protein disulphide isomerase — the main catalyst of oxidative protein folding in the endoplasmic reticulum. But yeast cells deficient in glutathione

biosynthesis have no defect in disulphide-bond formation, and reduced glutathione can even compete with reduced proteins for oxidizing equivalents^{10,11}. A protein called Ero1p, on the other hand, is associated with the membrane surrounding the endoplas-

mic reticulum, and is essential for the formation of disulphide bonds in yeast^{10,12}. So, Ero1p may be functionally equivalent to DsbB. However, the real source of oxidizing equivalents in the endoplasmic reticulum has not yet been identified, and additional components may also be required.

The extraordinary work by Bardwell and colleagues teaches us an important lesson. They were able to reconstitute the pathway of disulphide-bond formation by a single prerequisite — a pure protein preparation of DsbB. Or, as my first scientific mentor Wolfram Schäfer used to say, by "the secret of the pure substance".

Rudi Glockshuber is at the Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule, Hönggerberg, CH-8093 Zürich, Switzerland.

e-mail: rudi@mol.biol.ethz.ch

- Gilbert, H. F. *J. Biol. Chem.* **272**, 29399–29492 (1997).
- Bader, M., Muse, W., Ballou, D. P., Gassner, C. & Bardwell, J. C. A. *Cell* **98**, 217–227 (1999).
- Bardwell, J. C. A., McGovern, K. & Beckwith, J. *Cell* **67**, 581–589 (1991).
- Bardwell, J. C. A. *et al. Proc. Natl Acad. Sci. USA* **90**, 1038–1042 (1993).
- Bader, M., Muse, W., Zander, T. & Bardwell, J. C. A. *J. Biol. Chem.* **273**, 10302–10307 (1998).
- Kobayashi, T. & Ito, K. *EMBO J.* **18**, 1192–1198 (1999).
- Missiakas, D., Georgopoulos, C. & Raina, S. *EMBO J.* **13**, 2013–2020 (1994).
- Bessette, P. H. *et al. J. Biol. Chem.* **274**, 7784–7792 (1999).
- Rietsch, A., Belin, D., Martin, N. & Beckwith, J. *Proc. Natl Acad. Sci. USA* **93**, 13048–13053 (1996).
- Fränd, A. R. & Kaiser, C. A. *Mol. Cell* **1**, 171–182 (1998).
- Cuozzo, J. W. & Kaiser, C. A. *Nature Cell Biol.* **1**, 130–135 (1999).
- Pollard, M. G., Travers, K. J. & Weissman, J. S. *Mol. Cell* **1**, 171–182 (1998).

Earth science

Does global cooling reduce relief?

Eric Small

The nature of the interactions between climate and tectonic processes is strongly debated. It has been proposed that a shift towards a cooler and more erosive climate would enhance topographic relief in mountain ranges¹. The isostatic response to this change in relief (whereby parts of the Earth's lithosphere rise if material is removed and sink if it is deposited) could raise mountain crests even higher, driving climate to an even more erosive state and forming a positive feedback loop. On page 39 of this issue, Whipple *et al.*² present an alternative hypothesis — that the change to a more erosive climate would lead to either a decrease or no change in relief in active mountain ranges, indicating a negative or non-existent coupling between cooling and mountain uplift.

Sorting out the connection between climate and tectonics is important for understanding why the Earth has become progressively cooler over the past 50 million

years³. It has been proposed that uplift of the Himalayas, Tibetan plateau and other mountain ranges during the Cenozoic era (from 64 million years ago) would result in cooler global temperatures in three ways. First, persistent snow cover at higher altitudes would enhance the Earth's reflectivity. Second, increased uplift at mid-latitudes would dramatically alter atmospheric circulation, as demonstrated by global-climate models⁴. And third, enhanced weathering in mountainous regions would weaken the natural greenhouse effect, because weathering of silicate rocks consumes atmospheric CO₂.

Of course, uplift is only a reasonable source of Cenozoic cooling if it preceded the observed climate change. It is not disputed that tectonic forces uplifted the Himalayas and Tibetan plateau at some point during the Cenozoic. But much of the evidence for Cenozoic uplift in other parts of the world is ambiguous and may instead reflect global cooling¹. Greater erosion in mountainous

regions, accelerated deposition in sedimentary basins and cooler temperatures inferred from terrestrial fossils all may reflect the transition to a cooler, more erosive climate, rather than tectonic uplift. The geological record shows that cooling was accompanied by accelerated erosion, which may have been due to greater storminess related to a steeper equator-to-pole temperature gradient, and more widespread glaciation.

Even though uplift of mountains (other than those in central Asia) may not be the primary source of Cenozoic climate change, in 1990 Molnar and England¹ proposed that the response of mountain systems to global cooling may act as a positive feedback. They argued that if erosion is concentrated on valley floors in a cooler, glaciated climate, then relief in mountain ranges will increase and the isostatic response to this erosion would raise mountain peaks even higher, promoting further cooling (Fig. 1). The possibility of such a positive feedback mechanism introduces a 'chicken or egg' problem¹ — that is, which came first, global cooling or mountain uplift? The existence of the proposed feedback depends critically on whether or not the transition to a more erosive climate enhances relief in mountain ranges. The analysis of Whipple *et al.*² shows that in tectonically active mountains ranges a more erosive climate should reduce relief in fluvial landscapes and produce little change in relief in glaciated landscapes (Fig. 1).

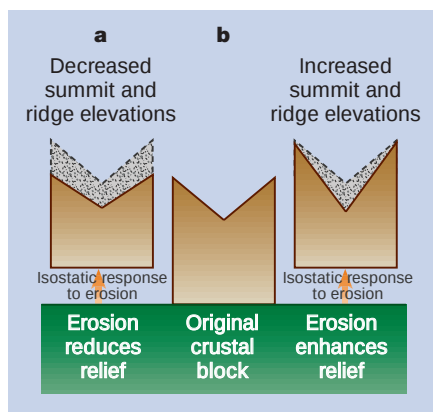


Figure 1 Does global cooling produce relief? Erosion of a mountain range, b, can lead to reduced, a, or enhanced, c, topographical relief (dashed line is topography before erosion, solid line is present topography). In either case, erosion causes the same isostatic uplift of the mountain range, but the height of the peaks following erosion differs. A new analysis by Whipple *et al.*² suggests that changes in erosion from global cooling should reduce relief, a, rather than increase relief, c, which could yield further cooling¹.

In non-glacial, tectonically active mountain ranges, the elevation drop along bedrock stream channels accounts for nearly all of the drainage basin relief. So, Whipple *et al.*² studied the response of bedrock stream channels to changes in climate. The authors

used the widely applied 'stream power erosion law'³, which states that the erosion rate of a bedrock channel is a power-law function of drainage area and channel gradient, and is scaled by the coefficient of erosion, k . Higher values of k represent a more erosive climate or less resistant rocks, among other things. The authors show that the relief along a bedrock channel varies inversely with k , assuming both steady-state conditions (where erosion balances rock uplift) and that uplift and k are spatially uniform across a mountain range. This means relief should be reduced in a more erosive climate compared with a less erosive climate. Although this is the opposite of what Molnar and England¹ proposed, it is analogous to the seemingly obvious notion that relief should be lower in a mountain range with more easily eroded rocks. But what happens in real mountain ranges? The analysis holds when the assumptions are relaxed, except for the case when k increases along most of a channel but decreases in the headwaters. It is difficult to assess the likelihood of this exception, as the relationship between k and climate is complex and poorly understood, making this a prime area for future research.

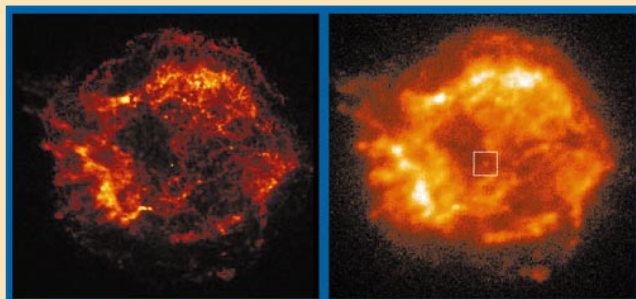
The response of river channels to climate change is only half the story. Increases in mountain relief and isostatic uplift of peaks may result from the onset of alpine glaciation, triggered once again by Cenozoic global cooling¹. Whipple *et al.* argue that erosion by

Astronomy

Chandra opens its eyes

After more than 20 years in the making, and one month after its launch, NASA's Chandra X-ray Observatory had its first look at the X-ray sky last week. One of the first images taken was of Cassiopeia A, the extremely hot remains of a massive star that exploded over 300 years ago. The Chandra image of Cassiopeia A is shown here on the left, whereas the image taken by an earlier X-ray satellite, ROSAT, is shown on the right. The highly detailed images produced by Chandra reveal a tantalizing bright spot near the centre, which could be the remains of the exploded star. Longer exposure images of Cassiopeia A from Chandra should help settle the question of whether the supernova left behind a neutron star, a black hole or nothing at all.

The world's newest and largest X-ray observatory has ten times greater resolution and is 50 to 100 times more sensitive than any



previous X-ray telescope. This means that Chandra — named after the late astrophysicist Subrahmanyan Chandrasekhar — will be able to see details as fine as most optical telescopes. In fact, over the next five to ten years NASA's latest observatory is expected to do for X-ray astronomy what the Hubble Space Telescope has done for optical astronomy.

Where Chandra will really come into its own is in the detection of faint X-ray sources. A bright source such as Cassiopeia A has already been imaged by radio, optical and

X-ray telescopes. But when Chandra turns its attention to the same piece of sky as observed by Hubble during its Deep Field survey, it should be able to see X-rays from hot, young stars in distant galaxies, which may help us to understand more about star formation in the early Universe.

Chandra is still in its early calibration stage, but this evidence that it is working successfully is a huge relief for X-ray astronomers and engineers, who have invested years in the project. Unlike the Hubble Space Telescope, Chandra

cannot be repaired in space if anything goes wrong. This is because it is in a much higher orbit than Hubble, which required several visits for emergency repairs by the Space Shuttle.

But whatever happens to Chandra, the outlook for X-ray astronomy is better than it's been for a long time. Chandra will be joined in the next six months by two more X-ray telescopes. The European Space Agency's X-ray Multi-Mirror satellite will follow a similar orbit to Chandra and will have a total collecting area four times that of Chandra's mirrors, although its resolution won't be as high. A third X-ray telescope will be launched by the Japanese early next year. The Astro-E telescope will orbit much lower than the other two, and will not have such a high resolution. Its task will be to collect ultrashort, 'hard' X-rays emitted by the most violent explosions and turbulent galaxies.

Sarah Tomlin

alpine glaciers generates relief in four ways: valley widening, lateral support for valley walls by glacial ice, and formation of hanging valleys and overdeepenings. On the basis of observations from several mountain ranges, the authors suggest that the relief created by each process should scale with ice thickness, which is typically several hundred metres for alpine glaciers. This hypothesis needs to be thoroughly tested. If it is valid, then the maximum relief produced by alpine glaciers could only be several hundred metres, yielding negligible uplift of mountain peaks⁶.

If Cenozoic cooling has not generated relief in mountains worldwide, as suggested by Whipple *et al.*², then the proposed positive feedback between cooling and uplift does not exist. And we have a solution to the chicken-or-egg problem: uplift of the Himalayas and Tibet may be an important source of Cenozoic cooling, but the topo-

graphic response to this climate change has not yielded additional cooling. In other words, the chicken's eggs may not be fertile. For this issue to be resolved definitively, new techniques are needed to unravel the history of mountain relief preserved in the geological record⁷. Only then will we know whether or not there is a rooster in the hen house. ■

Eric Small is in the Department of Earth and Environmental Science, New Mexico Tech., Socorro, New Mexico, 87801-4796, USA.

e-mail: esmall@nmt.edu

1. Molnar, P. & England, P. *Nature* **346**, 29–34 (1990).
2. Whipple, K. X., Kirby, E. & Brocklehurst, S. H. *Nature* **401**, 39–43 (1999).
3. Raymo, M. E. & Ruddiman, W. F. *Nature* **359**, 117–122 (1992).
4. Ruddiman, W. F. & Kutzbach, J. E. *J. Geophys. Res.* **94**, 18409–18427 (1989).
5. Howard, A. D., Dietrich, W. E. & Siedl, M. A. *J. Geophys. Res.* **99**, 13971–13986 (1994).
6. Small, E. E. & Anderson, R. S. *Geology* **26**, 123–136 (1998).
7. House, M. A., Wernicke, B. P. & Farley, K. A. *Nature* **396**, 66–69 (1998).

Apoptosis

Akt is more than just a Bad kinase

Asim Khwaja

The relentless search for the molecules and processes that control programmed cell death (apoptosis) reflects the biological importance of a phenomenon that is central to the smooth running of multicellular machines. Whether a cell lives — or dies — depends on external cues, such as the presence of soluble growth factors, contact with neighbouring cells or adherence to components of the extracellular matrix. These changes in the cell's environment are integrated into a survival response by what are often complex signal-transduction mechanisms. When things go wrong, the result can be malignant or degenerative disease.

A serine/threonine kinase known as Akt/protein kinase B has been identified as an important component of pro-survival signalling pathways (reviewed in ref. 1), and the race has been on to identify its substrates. A sheaf of papers, including reports by Ozes *et al.*² and Romashkova and Makarov³ in this issue (pages 82 and 86, respectively), now tell us more about how Akt exerts its anti-apoptotic effects. In particular, these studies show that Akt alters the activity of factors that regulate gene transcription.

The Akt kinase is activated via the phosphoinositide-3-OH kinase, PI(3)K, signalling pathway¹. It is dysregulated in a variety of tumours by several different mechanisms: by overexpression of the protein; by constitutive activation in tumours harbouring mutant Ras oncogenes; or by inactivation of an inhibitory phosphatase^{1,4}. At first, the search for Akt substrates centred on

cellular proteins with well-established roles in apoptosis. For example, phosphorylation of the pro-apoptotic Bad protein by Akt was

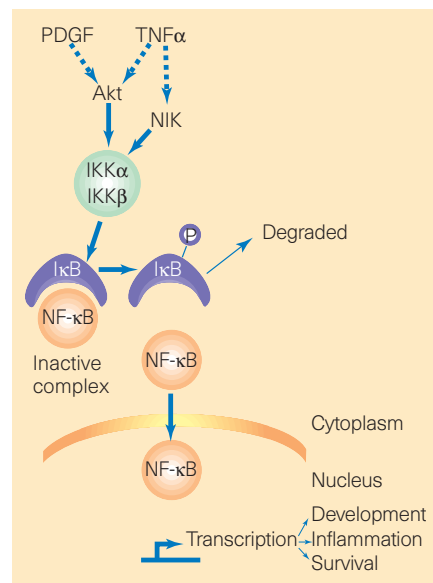


Figure 1 Cell-survival factor Akt is involved in activation of NF-κB, a regulator of gene transcription. Addition of tumour-necrosis factor-α (TNF-α) or platelet-derived growth factor (PDGF) results in activation of IκB kinases (IKKs), probably owing to cooperation between Akt and the NF-κB-inducing kinase (NIK). NF-κB is found in an inactive cytoplasmic complex with its inhibitor IκB. Phosphorylation of IκB by IKKs results in its degradation, allowing NF-κB to move to the nucleus and activate transcription.

shown⁵ to decrease apoptosis by preventing Bad from binding and inhibiting an anti-apoptotic protein called Bcl-x_L. Subsequently, Akt-induced phosphorylation of a death protease called caspase-9 was found⁶ to decrease apoptosis by inhibiting the protease activity directly.

As well as using post-translational mechanisms to change protein function directly, survival factors can act by altering gene transcription. The work of Ozes *et al.*² and Romashkova and Makarov³, as well as a report by Kane *et al.*⁷ in *Current Biology*, now links Akt to regulation of signalling via the nuclear factor-κB (NF-κB) transcription-factor pathway. The NF-κB protein promotes expression of anti-apoptotic genes⁸, so activation of NF-κB by Akt should decrease the likelihood of cell death. Normally, NF-κB is held in the cytoplasm as an inactive complex with its inhibitor IκB. After the cell is stimulated by pro-inflammatory cytokines — in particular, tumour-necrosis factor-α (TNF-α) and interleukin-1 — and also by certain growth factors, IκB is phosphorylated and targeted for degradation. This allows NF-κB to translocate to the nucleus, where it increases the transcription of genes concerned with development, inflammation, host defence and cell survival (Fig. 1).

Romashkova and Makarov³ show that platelet-derived growth factor (PDGF), a potent survival factor for fibroblasts, causes activation of NF-κB, and that this pathway depends on activation of Akt. They show a direct association between Akt and the kinases that phosphorylate IκB (the IKKs). Using dominant-negative proteins, the authors show that one IKK subunit, IKKβ, is involved in PDGF-induced activation of NF-κB. In their system, TNF-α-induced activation of NF-κB does not involve either PI(3)K or Akt.

By contrast, Ozes *et al.*², working with epithelial cells, show that TNF-α activates Akt, and that inhibition of Akt blocks the TNF-α-induced activation of NF-κB. Ozes and colleagues propose a mechanism for the effects of Akt. They show that Akt associates with, and phosphorylates, another IKK subunit, IKKα. When these authors introduced a non-phosphorylatable IKKα mutant, they blocked TNF-α-induced degradation of IκB, thereby inhibiting translocation of NF-κB to the nucleus. Moreover, Kane *et al.*, working with a T-cell line, find that activation of NF-κB is potentiated by Akt, most likely due to increased degradation of IκB. These findings not only tell us more about how Akt might control cell survival, but they also indicate an unexpected role for Akt in inflammation and immunity.

How do these findings fit with what is already known about Akt signalling? Earlier this year^{9,10}, members of the Forkhead family of transcription factors were identified as

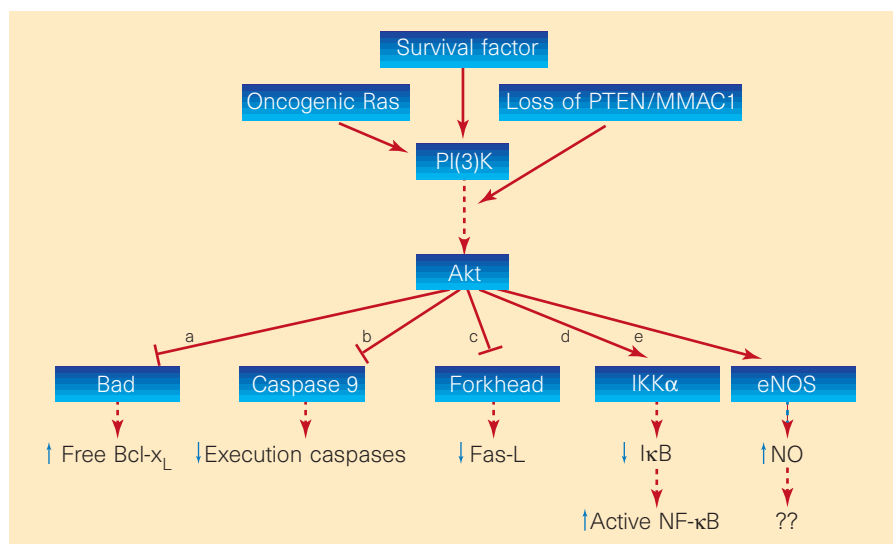


Figure 2 Potential mechanisms by which the Akt serine kinase can regulate cell survival. Phosphoinositide-3-OH kinase (PI(3)K) signalling can be activated by extracellular survival factors, by activating mutations in the Ras oncogene, or by a deletion/inactivating mutation of PTEN/MMAC1 (phosphatase and tensin homologue/mutated in multiple advanced cancers), a phospholipid phosphatase. This leads to activation of Akt. Phosphorylation by activated Akt can then: a, sequester the pro-apoptotic protein Bad, leading to increased free (anti-apoptotic) Bcl-x_L; b, inhibit the activity of the cell-death protease caspase-9, diminishing the activation of target execution caspases; c, prevent nuclear localization of Forkhead transcription factors, reducing the transcription of Fas ligand (Fas-L); d, promote the degradation of IκB and increase the activity of NF-κB, leading to increased synthesis of anti-apoptotic proteins; or e, increase the activity of endothelial nitric oxide synthase (eNOS), leading to increased production of nitric oxide (NO).

direct targets of Akt. Genetic studies in the nematode worm *Caenorhabditis elegans* had previously shown¹¹ that a PI(3)K/Akt pathway suppresses the function of DAF-16, a member of the Forkhead family. In the worm, this pathway regulates aspects of metabolism and longevity. Phosphorylation of mammalian members of the Forkhead family by Akt results in their being retained in the cytoplasm, preventing them from affecting transcription. In some cells this mechanism may regulate apoptosis by inhibiting the transcription of the gene that encodes the death-activating Fas ligand⁹ (Fig. 2). No doubt, though, other targets will also be involved.

Also this year, two groups^{12,13} have shown that Akt can influence the production of nitric oxide. It does so by directly phosphorylating nitric oxide synthase in the endothelium. Although the nitric oxide produced by this pathway is mainly involved in modulating vascular tone and the formation of new blood vessels, there is evidence¹⁴ linking it to pathways that regulate apoptosis. Endothelial cell-survival pathways are known to involve Akt, and it remains to be seen whether nitric oxide plays a part in this process.

The research on Akt signalling shows that a single regulatory molecule can exert its anti-apoptotic effects in a variety of ways. One implication of these results is that arm-chair-based generalizations on how apoptosis is controlled are no longer tenable. We

should not be surprised when our own favourite cell line behaves differently to someone else's — such cells will probably have developed individual molecular quirks that maintain their immortality. A challenge now is to make sense of the networks that control apoptosis, and the focus may need to shift away from cell lines to primary tissues. To find out what molecules are involved in controlling apoptosis in the physiological or disease process that you are interested in, you will have to head for the bench. This means that the huge research effort into the mechanics of cell death is likely to continue for some time.

Asim Khwaja is in the Department of Haematology, University College London Medical School, 98 Chenies Mews, London WC1E 6HX, UK. e-mail: a.khwaja@ucl.ac.uk

1. Downward, J. *Curr. Opin. Cell Biol.* **10**, 262–267 (1998).
2. Ozes, O. N. *et al. Nature* **401**, 82–85 (1999).
3. Romashkova, J. A. & Makarov, S. S. *Nature* **401**, 86–90 (1999).
4. Maehama, T. & Dixon, J. E. *Trends Cell Biol.* **9**, 125–128 (1999).
5. Datta, S. R. *et al. Cell* **91**, 231–241 (1997).
6. Cardone, M. H. *et al. Science* **282**, 1318–1321 (1998).
7. Kane, L. P., Shapiro, V. S., Stokoe, D. & Weiss, A. *Curr. Biol.* **9**, 601–604 (1999).
8. Wang, C. Y., Mayo, M. W., Korneluk, R. G., Goeddel, D. V. & Baldwin, A. S. Jr *Science* **281**, 1680–1683 (1998).
9. Brunet, A. *et al. Cell* **96**, 857–868 (1999).
10. Kops, G. J. *et al. Nature* **398**, 630–634 (1999).
11. Paradis, S. & Ruvkun, G. *Genes Dev.* **12**, 2488–2498 (1998).
12. Dimmiller, S. *et al. Nature* **399**, 601–605 (1999).
13. Fulton, D. *et al. Nature* **399**, 597–601 (1999).
14. Dimmiller, S., Haendeler, J., Nehls, M. & Zeiher, A. M. *J. Exp. Med.* **185**, 601–607 (1997).

Daedalus

Slip through the water

Daedalus once invented a red-hot boat. The water could never touch its hull, which was insulated by a thin layer of steam. So it had extremely low skin-friction, and no trouble with fouling weeds or barnacles either. But its fuel costs were high, and its insulation problems severe.

He now returns to the field with a converse scheme. He wants to coat a ship with ice. He argues that ice must have the lowest skin-friction against water of any solid. Furthermore, in its flow past the ice, the water would tend to melt away preferentially those features that obstructed the stream lines. The hull would shape itself into the perfect streamline form for minimum drag. But again, the idea suffers from heat-transfer problems. This time, Daedalus has a way out. He wants to coat his ship with foamed ice.

Foamed ice, like foamed polystyrene, would be an excellent thermal insulator. DREADCO engineers are now exploring ways of making it. The simplest idea is just to make the surface of the cooled hull porous, and pump air out of the holes as the water freezes onto it. Another is to saturate the surrounding water with a gas such as carbon dioxide, which is expelled by water as it freezes. It might even be possible to generate the gas electrolytically, using electrodes integral with the hull. But whichever way proves best, the ship will build up its primary ice-coating in port, and only have to maintain it against melting at sea. A good refrigerator can pump heat at several times its own power-drain. Daedalus estimates that a few tens of kilowatts of refrigeration could maintain a foamed ice-sheet over the wetted surface of even a sizeable freighter. Its slippery coating, proof against fouling organisms, should save many hundreds of kilowatts of propulsive power. And in a maritime accident, the ship could make instant rafts and lifeboats of foamed ice, to support the crew until help arrived.

Foamed ice should have many other uses too. The DREADCO team hopes to develop a generator to make it in quantity on land. A given volume of water would make a much larger volume of foamed ice. Unlike snow, it would have a closed-cell structure, impermeable and very rigid; and its low thermal conductivity would limit its rate of melting. It would be ideal for temporary structures such as crash barriers, fire breaks, scaffolding, ski-ramps and foamed igloos for emergency shelter.

David Jones

Why cars in the next lane seem to go faster

The temptation to change lanes on a motorway may be prompted by an illusion.

Switching lanes while driving along a busy road can be a risky manoeuvre. It is often instigated on the driver's judgement that the cars in the next lane are moving faster than those in the driver's own lane. But faulty intuition¹⁻³ may cause people to overestimate the speed of vehicles in the next lane, believing that they are moving faster even when both lanes have the same average speed. We suggest that this illusion occurs because more time is generally spent being overtaken (passed) by other vehicles than is spent in overtaking them. Knowing that this effect is illusory might encourage drivers to resist small temptations to change lanes.

We used computer simulations to create two lanes of traffic with identical characteristics, except that their congestion varied depending on random starting gaps. Vehicles then accelerated if no other vehicle was within the minimum headway distance, otherwise they decelerated. Minimum headway distances increased with higher velocities to prevent collisions.

We evaluated statistically the movements of an individual driver compared with all the others in the next lane. Under baseline conditions, many one-second epochs produced a change in the driver's relative position. Epochs in which the index vehicle was overtaken were more frequent than epochs in which the index vehicle was overtaking another vehicle (Fig. 1). The total number of vehicles that passed the driver was balanced by the total number of vehicles that were overtaken by the driver because of multiplicity in some epochs.

Doubling the acceleration generated a larger difference, and halving the minimum headway distance, to represent 'tailgating', greatly increased the difference. Reducing the frequency with which the index driver glanced at the next lane from once every second to once every two seconds attenuated the difference. No combination yielded a slower apparent speed for the next lane.

We also videotaped traffic sequences by mounting a camera in a moving vehicle and filming the side-view perspective of the next lane on a congested road. When a section of videotape showing a slightly slower average speed in the next lane was screened to driving students ($n = 120$), 70% stated that the next lane was moving faster and 65% said they would change lane if possible.

From these results, we suggest that drivers are responding to an illusion: namely, that the next lane on a congested road appears to be moving faster than the driver's present lane, even when both lanes have the same

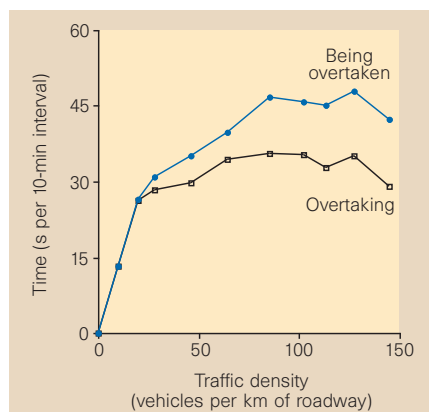


Figure 1 Relation of traffic density to the amount of time spent overtaking and being overtaken. All time estimates have standard errors of less than 5 seconds that decrease with traffic density. For example, a driver travelling for 10 minutes on a roadway with congestion of about 100 vehicles per kilometre would spend 47 seconds being overtaken by vehicles in the other lane and about 35 seconds overtaking vehicles in the other lane if both lanes have the same average speed. Details of simulation models are available from the authors.

average speed. This occurs because vehicles spread out when moving quickly and pack together when moving slowly. A driver can therefore overtake many vehicles in a brief time interval, but it takes much longer for the driver to be overtaken by the same vehicles.

Other aspects of human perception may accentuate the impression that the next lane is moving faster. Differential surveillance can occur because drivers look forwards rather than backwards, so vehicles that are overtaken become invisible very quickly, whereas vehicles that overtake the index driver remain conspicuous for much

longer⁴. Moreover, a driver is more likely to glance at the next lane for comparison when he is relatively idle⁵ while moving slowly.

Even if attention was not focused in particular directions and was evenly spaced in time, human psychology may make being overtaken (losing) seem more salient than the corresponding gains⁶. Furthermore, misconceptions about randomness can make runs of overtaking and being overtaken seem unduly prolonged^{7,8}. Our study highlights the effects of congestion and the increasing importance of the illusion, given that the number of miles travelled by vehicles is increasing at a much faster rate than the amount of roadway^{9,10}.

Donald A. Redelmeier*,

Robert J. Tibshirani†

*University of Toronto, Sunnybrook and Women's College Health Sciences Centre, G-151, 2075 Bayview Avenue, Toronto, Ontario M4N 3M5, Canada
e-mail: dar@ices.on.ca

†Departments of Health Research and Policy and of Statistics, Stanford University, Stanford, California 94305-5405, USA

1. Snowden, R. J., Stimpson, N. & Ruddle, R. A. *Nature* **392**, 450 (1998).
2. Walton, D. & Bathurst, J. *Accident Anal. Prevent.* **30**, 821-830 (1998).
3. Tversky, A. & Kahneman, D. Q. *J. Econ.* **107**, 1039-1061 (1991).
4. Larson R. C. *Operat. Res.* **35**, 895-904 (1987).
5. Angrilli, A., Cherubini, P., Pavese, A. & Manfredini, S. *Percept. Psychophys.* **59**, 972-982 (1997).
6. Tversky, A. & Kahneman, D. *Science* **185**, 1124-1131 (1974).
7. Gilovich, T., Vallone, B. & Tversky, A. *Cogn. Psychol.* **17**, 295-314 (1985).
8. Feller, W. *An Introduction to Probability Theory and its Applications* Vol II (Wiley, New York, 1966).
9. Helbing, D. & Huberman, B. A. *Nature* **396**, 738-740 (1998).
10. Bureau of the Census *Statistical Abstract of the United States 1996* 116th edn (Government Printing Office, Washington DC, 1997).

Rare-earth metals

Is gadolinium really ferromagnetic?

Gadolinium is accepted to be one of the four ferromagnetic elements, along with iron, cobalt and nickel, although its Curie point, T_C (the temperature above which ferromagnetism is lost), is only 292 K. Ferromagnets exhibit a characteristic divergence of their susceptibility at T_C , but no such divergence in any direction is found for needle-shaped crystals of gadolinium. Instead, the susceptibility diverges at a lower spin-reorientation temperature, T_{sr} , of 225 K, where the anisotropy changes sign. We propose that the magnetic order between

T_{sr} and T_C is not truly ferromagnetic, but is akin to the incommensurate order found in erbium.

It used to be thought¹ that gadolinium had a helical spin structure similar to that of terbium, dysprosium and holmium, but that it became ferromagnetic when a small field (~ 1 kA m⁻¹) was applied. This idea was discounted after neutron diffraction² showed that the turn angle must be smaller than 2 degrees. Gadolinium is considered to be the only simple ferromagnet among the rare-earth metals. Moments lie along the hexagonal c -axis between T_C and T_{sr} , where they become inclined at an angle to c which reaches 65 degrees at 180 K.

Previous studies³ of the anisotropic susceptibility, $\chi = M/H$, where M is the

magnetization and H is the magnetic field strength, have been performed on crystals with a demagnetizing factor, N , of 0.1 to 0.3. The demagnetizing factor depends only on shape and determines the strength of the reverse field that appears inside a sample when it is magnetized; χ is limited by the demagnetizing effect to $1/N$. We therefore use long, thin samples with $N \approx 0.01$, and measure χ in a small alternating-current field of 8 A m^{-1} , compensating for the Earth's field with a pair of Helmholtz coils.

Figure 1a shows that the real component $\chi'(T)$ measured perpendicular to the c -axis has a sharp peak at the magnetic transition with a maximum value of 49 ± 4 . As the temperature decreases further, χ' increases to saturate near the demagnetizing-limited value $1/N = 77$. When the field is parallel to c , there is only a shoulder in $\chi'(T)$ at T_C and the value does not exceed 10. $\chi''(T)$ increases at lower temperature, reaching the demagnetizing limit $1/N = 66$ at 230 K.

Figure 1b shows the real and imaginary components, χ' and χ'' , for the two directions in the vicinity of T_C , after correcting for the demagnetizing effect⁴. The $\chi_{\perp}'(T)$ peak at the magnetic transition is accompanied by a small peak in the $\chi_{\perp}''(T)$ component, but $\chi_{\parallel}'(T)$ barely exceeds 10 at T_C and $\chi_{\parallel}''(T)$ is zero within experimental error. There is no evident divergence at T_C . To see whether the observed non-divergence of the c -axis susceptibility could be related to pinning of domain walls, we applied a 50-Hz alternating-current bias field of 80 A m^{-1} to mobilize the walls. The form of χ' and χ'' did not change.

The susceptibility of a ferromagnet normally diverges as $|\epsilon|^{-\gamma}$, where $\epsilon = (T - T_C)/T_C$ and the critical exponent γ is 1.38 for isotropic spins (Heisenberg model) or 1.24 for extreme uniaxial anisotropy (Ising model). The presence of uniaxial dipolar anisotropy⁵ should not prevent divergence at T_C .

Internal checks on the anomalous behaviour of $\chi_{\parallel}'(T)$ are provided by the divergence at lower temperatures in both directions and by the peak in $\chi_{\perp}(T)$. The susceptibility effectively reaches the demagnetizing-limited values at the spin-reorientation transition, but not at the Curie point. The sharp peak observed at T_C when the field is applied in the basal plane resembles the peak at the Néel point (at which anti-ferromagnetic materials become paramagnetic), T_N , of 230 K for terbium, where an in-plane helical structure is established⁶ and χ' reaches a value of 8. When magnetized in-plane near T_C , where magnetocrystalline anisotropy is negligible, the susceptibility of gadolinium resembles that of a long-period helix rather than a true ferromagnet. When magnetized along c ,

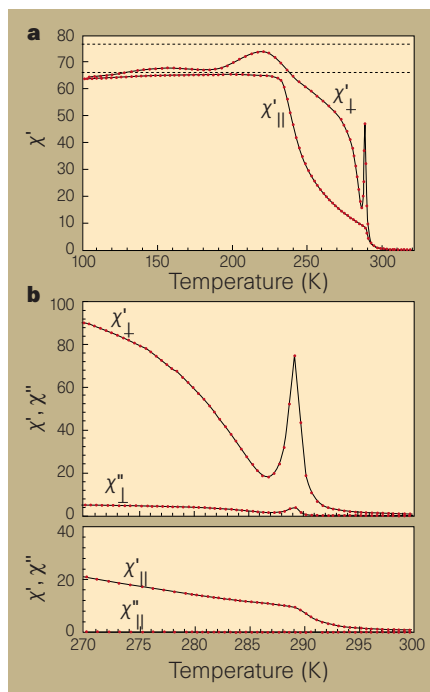


Figure 1 Magnetic susceptibility of a gadolinium crystal. **a**, Real and imaginary parts of the susceptibility measured with an alternating-current technique on long samples cut from a crystal of gadolinium with the field applied parallel ($\parallel$) and perpendicular ($\perp$) to the c -axis. The susceptibility is independent of frequency in the range 10 Hz to 1 kHz. The horizontal dotted lines indicate the saturation values expected for a divergent susceptibility, given the sample shape. **b**, Internal susceptibility of gadolinium in the vicinity of T_C , corrected for the demagnetizing effect caused by the sample shape.

there is a long-period sinusoidally modulated structure like that of erbium.

The value of the exchange stabilization energy of the helix, $J(q_m) - J(0)$, may be estimated from the peak susceptibility using $\chi_{\max} = \mu_0 N m^2 / \{S^2 [J(q_m) - J(0)]\}$ (ref. 7), where μ_0 is the permeability of free space, m is the atomic moment of gadolinium, $S = 7/2$, J is the exchange energy and q_m is the reduced wave vector in the c -direction; $\chi_{\max} = 75$ gives $J(q_m) - J(0) = 0.013 \text{ K}$. Extrapolating from the heavy rare-earth metals, where $J(q_m) - J(0) \approx q_m^4$, we estimate that $q_m = 0.029 \text{ \AA}^{-1}$; the corresponding wavelength, λ_m , is $70 \text{ \AA}$.

Examination of gadolinium close to T_C , using techniques to observe the long-period modulation directly, may confirm that gadolinium is not really ferromagnetic.

J. M. D. Coey, V. Skumryev, K. Gallagher

Physics Department, Trinity College, Dublin 2, Ireland

1. Belov, K. P. & Pedko, A. L. *Zh. Eksp. Teor. Fiz.* **47**, 87 (1962). (English transl., *Sov. Phys. JETP* **15**, 62–64 (1962).)
2. Cable, J. W. & Wollan, E. O. *Phys. Rev.* **165**, 733–734 (1968).
3. Dan'kov, S. Y., Tishin, A. M., Pecharsky, V. K. & Gschneider, K. A. Jr *Phys. Rev. B* **57**, 3478–3489 (1998).
4. Chen, D.-X. & Sanchez, A. J. *Appl. Phys.* **70**, 5463–5477 (1991).
5. Geldart, D. J. W., Hargraves, P., Fujiki, N. M. & Dunlap, R. A. *Phys. Rev. Lett.* **62**, 2728–2731 (1989).
6. del Moral, A. & Lee, E. W. J. *Phys. F* **4**, 280–290 (1974).
7. Kitano, Y. & Nagamia, T. *Prog. Theor. Phys.* **31**, 1–37 (1964).

Physiology

Warm feet promote the rapid onset of sleep

Even healthy people occasionally have difficulty falling asleep. Psychological relaxation techniques, hot baths, soothing infusions of plant extracts, melatonin and conventional hypnotics are all invoked in the search for a good night's sleep. Here we show that the degree of dilation of blood vessels in the skin of the hands and feet, which increases heat loss at these extremities, is the best physiological predictor for the rapid onset of sleep. Our findings provide further insight into the thermoregulatory cascade of events that precede the initiation of sleep¹.

Our analysis combines data from two intervention studies designed to induce phase shifts in the circadian pacemaker^{2,3}. Healthy young men were given melatonin, bright light, or both, in the evening ($n = 8$), or a large carbohydrate-rich meal in either the morning or the evening ($n = 10$). These manipulations had different thermoregulatory effects and so gave a broad variance, enabling us to extract the best predictor variables for the latency of sleep onset.

Subjects lay in bed under controlled conditions (lighting less than 8 lux; 22°C), taking small snacks of constant caloric value every hour during wakefulness. The latency of sleep onset was defined as the time between lights out (24:00) and the first occurrence of stage 2 in the sleep EEG recordings. Heart rate, core body temperature (rectal) and proximal (combined infraclavicular, thigh, stomach, forehead) and distal (combined hands and feet) skin temperatures were continuously measured⁴ (and later collapsed into 30-min bins). Sleepiness was rated² every 30 min and saliva was collected for melatonin assay⁵. We calculated the distal-proximal temperature gradient (DPG), a measure of blood flow in distal skin regions (efficiently regulated by arteriovenous anastomoses)⁶ that provides an indirect index of distal heat loss.

By using a multiple regression model for repeated measures, with the latency of sleep onset as the dependent variable ($n = 18$ subjects, 144 data points; between-subjects differences taken into account), we found the highest correlation with the DPG averaged over the three data points between 22:30 and 24:00 ($r = -0.47$, $P < 0.0001$). Thus, the greater the distal vasodilation in the late evening, the shorter was the time taken to fall asleep. We found that correlations were weaker with subjective sleepiness ratings ($r = -0.33$, $P < 0.0001$), core body temperature ($r = 0.26$, $P < 0.005$), melatonin ($r = -0.15$, n.s.; after excluding the melatonin intervention data), rate of change of core body temperature ($r = -0.05$, n.s.) and heart rate ($r = 0.05$, n.s.).

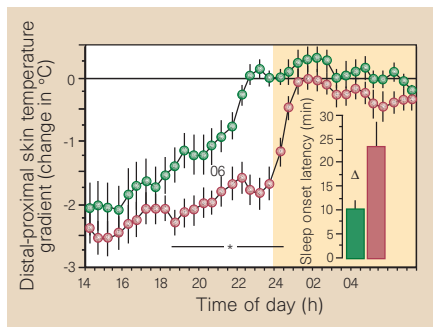


Figure 1 Time course of the distal-proximal skin-temperature gradient (DPG). The gradient is shown from 14:00 to 7:30 for observations with the most negative DPG values between 22:30 and 24:00 (large vasoconstriction before lights out at 24:00; pink symbols), compared with the time course of the most positive DPG values (large vasodilation before lights out; green symbols) (mean $\pm$ s.e.m., $n = 18$; asterisk indicates significant differences between data points; $P < 0.05$, Bonferroni-adjusted least significant differences). The shaded area indicates the lights-out period. These two extreme patterns were selected after the mean of the three DPG values between 22:30 and lights out at 24:00 had been rank ordered out of 8 observations for each subject. Sleep onset latency (inset) is significantly shorter when subjects were most vasodilated (green bar) before lights out (triangle indicates significant differences; paired t -test, $P < 0.001$).

In a backward stepwise regression analysis among all predictor variables, only DPG contributed significantly to the model; that is, vasodilation of distal skin regions was the best predictor of the body's readiness for sleep (Fig. 1). Because interventions such as light or large carbohydrate-rich meals differentially manipulated the independent variables, the effect on the dependent variable showed that the link between distal vasodilation and the ability to fall asleep is functional, not just correlative.

The circadian clock prepares the thermoregulatory system for vasodilation to begin in the early evening as sleepiness increases, followed by a drop in core body temperature. Even lying down increases sleepiness by redistributing heat in the body from the core to the periphery⁷. Turning out the light is a complex cognitive and physiological signal that also leads to vasodilation⁴. There is a tight correlation between the timing of the endogenous increase in melatonin in the evening and vasodilation, an effect that is mimicked by pharmacological doses of melatonin^{4,7}. Before bedtime, then, many overlapping events orchestrate the thermoregulatory overtone.

We would predict that classical hypnotics⁸ and other sleep-inducing aids all cause dilation of distal blood vessels and heat loss before the onset of sleep. A hot-water bottle at the feet, while not acting on mechanisms in the central nervous system that underly the regulation of sleep, can rapidly induce vasodilation. The resulting heat loss is most efficient when the ambient temperature is cool⁹. Some sleep disorders (particularly those associated with ageing

and somatic illness¹⁰) may be secondary to an inability to vasodilate and prepare the body for sleep.

Kurt Kräuchi, Christian Cajochen,

Esther Werth, Anna Wirz-Justice

Chronobiology and Sleep Laboratory,

Psychiatric University Clinic,

Wilhelm Klein-Strasse 27, 4025 Basel, Switzerland

e-mail: kurt.kraeuchi@pukbasel.ch

1. Campbell, S. S. & Broughton, R. J. *Chronobiol. Int.* **11**, 126–131 (1994).
2. Cajochen, C., Kräuchi, K., Danilenko, K. V. & Wirz-Justice, A. *J. Sleep Res.* **7**, 145–157 (1998).
3. Wirz-Justice, A. *et al.* *J. Sleep Res.* **7** (suppl. 2), 308 (1998).
4. Kräuchi, K., Cajochen, C., Möri, D., Hetsch, C. & Wirz-Justice, A. *Am. J. Physiol.* **272**, R1178–R1188 (1997).
5. Weber, J. M., Unger, I., Wirz-Justice, A. & Schwander, J. C. *J. Sleep Res.* **7** (suppl. 2), 302 (1998).
6. Rubinstein, E. H. & Sessler, D. I. *Anesthesiology* **73**, 541–545 (1990).
7. Kräuchi, K., Cajochen, C. & Wirz-Justice, A. in *Circadian Clocks and Entrainment* (eds Honma, K. I. & Honma, S.) 131–146 (Hokkaido Univ. Press, Sapporo, 1998).
8. Gilbert, S. S., van den Heuvel, C. J. & Dawson, D. J. *Physiol. (Lond.)* **514**, 905–914 (1999).
9. Aschoff, J. *Wien. Med. Woch.* **19/20**, 404–409 (1958).
10. Foley, D. J. *et al.* *Sleep* **18**, 425–432 (1995).

Biological motors

Connecting stalks in V-type ATPase

In all organisms, adenosine triphosphate (ATP) provides metabolic energy for driving energy-dependent processes. It is synthesized and/or utilized by enzymes known as F-type and V-type ATPases, which are small rotary motors^{1,2}. Both types consist of a headpiece, F_1 or V_1 , respectively, which is connected by a stalk region to the membrane-bound part, F_0 or V_0 . Electron microscopic analysis of negatively stained particles has revealed a peripheral stalk, or stator, between V_1 and V_0 of the V-type (Na^+)ATPase of the thermophilic bacterium *Clostridium fervidus*^{3,4}, like that in F-type ATPases^{5,6}. We have analysed many more particles and now present a more complete structure of the V-type ATPase stator moiety.

A central stalk in the ATPase rotates within a ring of three α - and three β -subunits in F_1 (refs 7, 8), or three A and three B subunits in V_1 , in discrete steps of 120°. At the F_0/V_0 end, the central stalk is connected to a ring of c subunits in the membrane. These subunits rotate against other subunits of F_0 and V_0 , allowing ion translocation at the interface². A stator structure in the form of additional connections between F_1/V_1 and F_0/V_0 must presumably be present to prevent futile rotation of the $\alpha\beta_3$ and A3B3 headpiece².

We classified a set of 7,500 molecular projections of detergent-solubilized, negatively stained V_1V_0 , and found a few preferential orientations in which V_1 has either two or three lobes. In about 28% of all

views, there are three connections between V_1 and V_0 , but in the rest either one (33%) or both (39%) of the peripheral connections appear to be missing. The three connections are well separated in the bilobed views (Fig. 1a). In the classes with trilobed views, the left connection is less well resolved because it is closer to the central stalk and overlaps more with V_1 (Fig. 1b). The two peripheral connections are each attached to an oval-shaped area of density on top of V_1 . In some other classes (not shown), one of the oval densities is absent, which correlates with the loss of a peripheral connection. We conclude that the intact V-type ATPase has a central stalk and two stator connections (Fig. 1c, d). Likely candidates for the stators are the bacterial V-type subunits I, E and F, several copies of which are present.

The discovery of the second stator raises several points. First, although it has been suggested that there may be two stators⁹, this feature has not been observed before. This could be because the intact structure is easily damaged on preparation, which can be detected only by classifying large sets of projections. The two stators may be unique for V-type ATPases because, in the stalk region of F-type ATPases, only the b subunit is present in two copies, as a dimer¹⁰, participating in one stator. It is harder to visualize the stator(s) in F-type ATPase by electron microscopy because the stalk region is much shorter than in V-type ATPase⁸.

Second, there could be more than two stators: three would match the three-fold symmetrical ring of the six large subunits

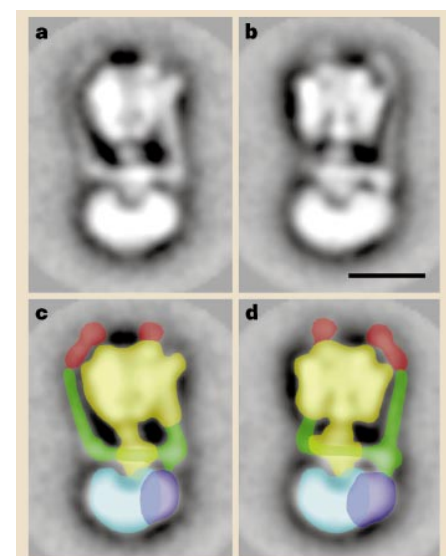


Figure 1 Electron microscopy images of V_1V_0 in side view. **a**, **b**, Views obtained by classification: **a**, bilobed view³; **b**, trilobed view. **c**, **d**, Model of the arrangement of the stator moiety (green) and its attachment to the V_1 headpiece (yellow) by the two oval densities (red). The view in **d** shows a larger additional density (dark blue) of V_0 on the right side than that in **c**. The stator moiety is attached to this additional density, which represents the static part of V_0 . The view in **d** is obtained by rotating **c** about 30° backwards on the left. Scale bar, 100 Å.

in V_1 . There is no evidence for this, but the presence of another stator connection is a possibility if there is overlap in projection.

Finally, it is not clear why the ATPase motor has such a complicated stator moiety, with an overall U-shaped form that must avoid friction with the central rotating stalk, although it is likely to be related to its mechanical stability.

**E. J. Boekema, J. F. L. van Breemen,
A. Brisson, T. Ubbink-Kok, W. N. Konings,
J. S. Lolkema**

Groningen Biomolecular Sciences and Biotechnology
Institute, University of Groningen,
9747 Groningen, The Netherlands
e-mail: boekema@chem.rug.nl

- Boyer, P. D. *FASEB J.* **9**, 559–561 (1995).
- Junge, W., Lill, H. & Engelbrecht, S. *Trends Biochem. Sci.* **22**, 420–423 (1997).
- Boekema, E. J., Ubbink-Kok, T., Lolkema, J. S., Brisson, A. & Konings, W. N. *Proc. Natl Acad. Sci. USA* **94**, 14291–14293 (1997).
- Boekema, E. J., Ubbink-Kok, T., Lolkema, J. S., Brisson, A. & Konings, W. N. *Photosynth. Res.* **57**, 267–273 (1998).
- Wilkens, S. & Capaldi, R. A. *Nature* **393**, 29 (1998).
- Böttcher, B., Schwarz, L. & Gräber, P. *J. Mol. Biol.* **281**, 757–762 (1998).
- Yasuda, R., Noji, H., Kinosita, K. & Yoshida, M. *Cell* **93**, 1117–1124 (1998).
- Abrahams, J. P., Leslie, A. G. W., Lutter, R. & Walker, J. E. *Nature* **370**, 621–628 (1994).
- Dschida, W. J. & Bowman, B. J. *J. Biol. Chem.* **267**, 18783–18789 (1992).
- Dunn, S. D. & Chandler, J. *J. Biol. Chem.* **273**, 8646–8651 (1998).

Insect behaviour

Evolutionary origins of bee dances

Although bumble-bees are highly social insects, their foraging has been considered to be managed as an individual initiative^{1–4}, in which each bumble-bee visits flowers not only to collect food, but also to gather information about other potential food sources⁵. Here we show that bumble-bees instead use a primitive, but surprisingly efficient, recruitment system: by performing extended excitatory runs in the nest, a single successful forager can alert the entire foraging force of a bumble-bee colony. But in contrast to what happens in other social bees, such as honeybees, the recruits are not informed about the location of the food. Instead, the successful forager brings home the odour of the newly discovered food source, conveying to the recruits information about the species of flower. These findings about bumble-bee communication shed new light on the early evolutionary origins of the elaborate dance language of the honeybee.

To investigate whether bumble-bees (*Bombus terrestris*) can communicate information about the discovery of a food source, we connected a nest box with a bipartite flight arena. A single forager was allowed to

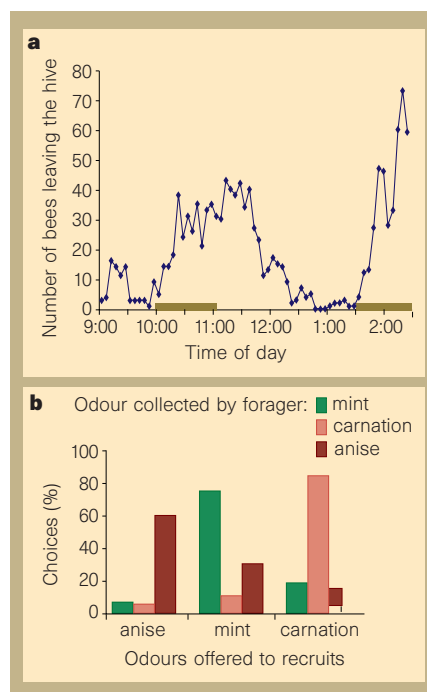


Figure 1 Recruitment in bumble-bees. **a**, The number of bees that leave the hive in a 5-minute period increases dramatically when one bee forages successfully (brown bars). **b**, Most bees choose the odour that is brought into the nest by a forager.

collect sucrose solution from an artificial flower in one half of the arena, whereas all other bees had access only to the other half of the arena, which did not contain food. This procedure ensured that interactions between bees could take place only in the nest, not at the food source. The number of bees entering the empty flight arena to search for food was counted during 12 periods of 1 hour each, when foragers were rewarded.

When compared with unrewarded control periods, the searching activity strongly increased when a single bee foraged ($P < 0.01$, Wilcoxon test; Fig. 1a). Successful foragers in the nest, instead of just emptying their crop load and continuing to forage, would typically spend several minutes running across the nest, frequently bumping into nestmates and occasionally buzzing their wings.

Odour-preference tests indicated that bees leaving the hive strongly preferred the odour that was brought home by the forager ($n = 90$, $P < 0.001$, χ^2 test; Fig. 1b). Because floral scents are species specific, the odour helps recruits to find the food source used by the successful forager. To test whether positional information is also conveyed, we trained three foragers to collect sucrose from a feeder positioned 100 m west of the hive in an open field. Recruits had the choice between that feeder and two additional feeders placed 100 m north and south of the nest. New recruits distributed themselves randomly at these three feeders ($P = 0.21$; χ^2 test). This failure of bumble-bees to recruit to specific points in space may explain why

their communication has previously been overlooked. Bumble-bees therefore have a recruitment system that is specific for a flower species but independent of its location. This system for transmitting information resembles the round dance used by honeybees when food is found in the immediate vicinity of the hive^{1,6}.

But why have bumble-bees not evolved a communication system that includes information about where to forage? Such a system can be very costly. Honeybee recruits may take more than an hour to decide where to go, even if only two different locations are advertised⁷. They also take a long time to find a food source after receiving information about its location⁸.

Bumble-bees live in small colonies, mainly in temperate habitats where floral food is less clumped than in the tropical habitats where the communication of honeybees and stingless bees evolved^{1,5,6}. The advantage of communicating location might therefore not offset the cost. It may be sufficient to specify flower species to recruits, who can then rely on their own memory or searching ability to find flowers but still know when and for what to forage.

Understanding the communication of bumble-bees is central to reconstructing the evolutionary origins of the honeybees' waggle dance, which is one of the most complex systems in animal communication. Such a reconstruction requires us to compare the honeybees' behaviour with that of their extant relatives, the bumble-bees and stingless bees, which have excitatory motor patterns that serve to recruit nestmates to food sources⁹. This behaviour may be derived from a social facilitation of activity at the nest entrance, which is widespread among social insects¹⁰. Bumble-bees also share with stingless bees⁹ and honeybees^{1,6} the ability to learn the floral odours brought back by returning foragers. The bumble-bee recruitment system might therefore resemble that of the last common ancestor of the eusocial bees.

A. Dornhaus, L. Chittka

Lehrstuhl für Zoologie II, Biozentrum,
Universität Würzburg, Am Hubland,
97074 Würzburg, Germany

e-mail: dornhaus@biozentrum.uni-wuerzburg.de

- Frisch, K. v. *Tanzsprache und Orientierung der Bienen* (Springer, Heidelberg, 1965).
- Esch, H. *Sci. Am.* **217**, 97–104 (1977).
- Dukas, R. & Real, L. A. *Oecologia* **94**, 244–246 (1993).
- Kirchner, W. H. & Towne, W. F. *Sci. Am.* **270**, 74–81 (1994).
- Heinrich, B. *Bumblebee Economics* (Harvard Univ. Press, Cambridge, Massachusetts, 1979).
- Seeley, T. *The Wisdom of the Hive: The Social Physiology of Honey Bee Colonies* (Harvard Univ. Press, Cambridge, Massachusetts, 1995).
- Michelsen, A., Lindauer, M. & Rohrseitz, K. in *27th Göttingen Neurobiology Conference* (eds Elsner, N. & Eysel, U.) 258 (Thieme, Göttingen, 1999).
- Wenner, A. M. & Wells, P. H. *Anatomy of a Controversy* (Columbia Univ. Press, New York, 1990).
- Lindauer, M. & Kerr, W. E. *Bee World* **41**, 29–71 (1960).
- Blackith, R. E. *Physiol. Comp. Oecol.* **4**, 388–402 (1957).

Geomorphic limits to climate-induced increases in topographic relief

Kelin X Whipple, Eric Kirby & Simon H. Brocklehurst

Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Recognition of the potential for strong dynamic coupling between atmospheric and tectonic processes has sparked intense cross-disciplinary investigation and debate on the question of whether tectonics have driven long-term climate change or vice versa. It has been proposed that climate change might have driven the uplift of mountain summits through an isostatic response to valley incision. Because isostasy acts to compensate mean elevations, the debate hinges on the question of whether climate change can significantly increase topographic relief or, more precisely, increase the volume of 'missing mass' between summits and ridges. Here we show that, in tectonically active mountain ranges, geomorphic constraints allow only a relatively small increase in topographic relief in response to climate change. Thus, although climate change may cause significant increases in denudation rates, potentially establishing an important feedback between surficial and crustal processes, neither fluvial nor glacial erosion is likely to induce significant isostatic peak uplift.

There is increasing recognition that geodynamic and geothermal responses to surface erosion and relief production provide a direct linkage between climate-driven denudation and tectonic processes^{1,2}. The coupled effects of topographic relief evolution and erosional unloading importantly influence the following: (1) isostatically compensated mountain summit elevations^{3,4} and thus interpretation of landscape morphology in terms of tectonic surface uplift^{5,6}; (2) spatial and temporal partitioning of tectonic and isostatic rock uplift, with potential direct feedback loops between zones of focused denudation and relief production and zones of focused crustal strain^{1,2}; and (3) near-surface geothermal gradients and thus interpretation of thermochronological data in terms of rock exhumation rates^{7,8}. The evolution of relief in response to climate change in particular is important in the debate over potential linkages between late Cenozoic uplift of the Himalayas and the Tibetan plateau and the onset of Quaternary glaciation^{3,5,9}.

Molnar and England⁵ argued that evidence cited for accelerated rates of Quaternary tectonism might in fact be produced by enhanced rates of valley incision induced by Quaternary climate change to stormier and/or glaciated conditions. The crux of the argument and central to the ensuing debate is the hypothesis that Quaternary climate change not only accelerated erosion but also caused a significant increase in topographic relief^{3,6,10,11}. Although some supporting data have been obtained^{6,10}, no convincing empiri-

cal evidence for significant climate-induced Quaternary peak uplift has been produced. Moreover, no comprehensive analysis of quantitative limitations to relief production in either glacial or non-glacial environments has hitherto been presented.

Here we provide a quantitative overview of the processes of erosion and relief production. We specifically focus on the question of the amount of relief production (or reduction) that can be expected from a change to a more erosive climate. We consider in turn the effects of an increase in fluvial erosivity, and the effects of a transition from fluvial to glacial erosion. Where current understanding of the erosion processes is insufficient to provide quantitative constraints (for example, glacial erosion), we cite empirical evidence and attempt to outline critical outstanding problems. We limit our focus to tectonically active orogens, intentionally avoiding the well known, limiting case of climate-induced incision into an undissected, elevated plateau^{4,5}.

Scales of relief

As with most morphometric parameters, relief varies with the scale of measurement, generally increasing with scale up to approximately the half-width of the mountain range. Thus it is necessary first to define measures of relief pertinent to the problem of isostatic response to denudation. Although relief varies over a continuum of scales, it is useful to define three fundamental components of relief: hillslope relief, tributary channel relief (elevation drop along tributaries), and trunk channel relief (elevation drop along the trunk channel) (Fig. 1). The sum of hillslope relief and trunk channel relief is equivalent to the often-used "drainage basin relief". Similarly, the sum of hillslope relief and tributary channel relief is roughly analogous to the commonly used measure of ridge line to valley bottom relief^{3,4,12,13}. In the case of non-glacial landscapes, channel relief must be further subdivided into colluvial, bedrock, and alluvial channel components (Fig. 1), owing to differences in operative erosion processes in each¹⁴. Colluvial channels are those channels where erosion is dominantly by debris-flow scour^{15,16}. In the case of glaciated landscapes, channel relief is simply the elevation drop on glaciated valley floors plus the elevation drop on bedrock and alluvial channels below the glacial limit.

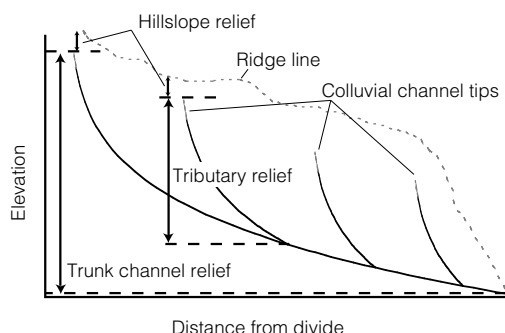


Figure 1 Sketch showing definitions of the components of fluvial relief. Heavy solid lines show trunk and tributary channel profiles projected onto a vertical plane through the axis of the drainage basin. Grey solid lines show debris-flow dominated or 'colluvial' channel tips. The light dashed line is a schematic representation of ridge line elevations projected onto the same plane as the channel profiles. Heavy dashed lines are horizontal reference lines used to define the various (labelled) components of relief. Downstream alluvial channel reaches are not shown.

Relief production in fluvial landscapes

Hillslope relief in tectonically active orogens is dictated by drainage density, and therefore hillslope lengths¹⁷, and rock mass strength^{18,19}. Here "tectonically active" can be defined as a condition in which rates of rock uplift, and commensurate channel incision, are greater

than soil production rates²⁰ such that slopes are stripped of soil, and hillslope denudation is dominated by deep-seated landsliding^{19,21,22}. Under these conditions, frequency distributions of hillslope gradients rapidly attain a maximum or saturation state beyond which they are statistically invariant¹⁹ (Table 1). With the exception of brief transients in which hillslopes may be temporarily oversteepened by rapid incision, hillslope relief in active orogens can only be increased by a decrease in drainage density. However, Schmidt and Montgomery¹⁸ have shown that longer hillslopes are less stable, and therefore hillslope relief itself probably reaches a maximum. Thus, a change of climate to a more erosive fluvial environment cannot significantly increase the hillslope component of relief in a tectonically active orogen³. Indeed, increasing the erosivity of the fluvial system is likely to increase drainage density, producing a commensurate decrease in hillslope relief²³.

Relief evolution in fluvial environments is most strongly dependent on the response of the bedrock channel system. Alluvial channels have gentle gradients, typically occur only in the foreland and foothills of active mountain ranges, and thus contribute little to overall relief. In addition, although the controls on the extent of colluvial channels are not well known, evidence from tectonically active non-glacial mountain ranges suggests that the elevation drop on bedrock channels typically comprises 80–90% of drainage basin relief (Fig. 1; Table 2). Therefore, with the caveat that the dynamics and extent of colluvial channels are at present poorly known²⁴, an analysis of the response of bedrock channel relief should represent the first-order characteristics of the evolution of relief in an unglaciated landscape.

Although many important questions remain unanswered in the study of bedrock channel erosion^{25,26}, the well-known “stream power erosion law” appears sufficiently robust to characterize the response of the bedrock channel to a change in the erosivity of the fluvial system. According to the stream power erosion law, local erosion rate is a power-law function of upstream drainage area (that is, $A = k_a x^h$ where k_a is a dimensional constant, x is distance downstream, and h is the reciprocal of the Hack exponent²⁷) and channel gradient ($S = -dz/dx$), such that a river profile evolution equation can be written as²⁸:

$$\frac{dz}{dt} = U(x, t) - K k_a^m x^{hm} \left| \frac{dz}{dx} \right|^n \quad (1)$$

where dz/dt is the rate of change of bed elevation, U is rock uplift rate (defined relative to base-level), K is a dimensional coefficient of erosion, and m and n are positive constants that reflect erosion processes, basin hydrology, and channel hydraulic geometry^{26,28}.

The coefficient of erosion K is influenced by many climate-related factors including precipitation, storminess, sediment flux, and channel width²⁶. The quantitative relationship between K and climate variables is complex, as these factors may adjust differently in response to a given climate change. We do not address this unresolved problem. Rather we ask the simpler question: what is the effect of a hypothetical shift to a ‘more-erosive’ (higher- K) condition? Further, we first treat the simplified case of a uniform change in K and then briefly consider the potential effects of non-uniform changes in erosivity.

Under steady-state conditions (for uniform K and U) where erosion balances rock uplift ($dz/dt = 0$), equilibrium channel gradient is a power function of drainage area (A):

$$S = k_s A^{-\theta} \quad (2a)$$

$$k_s = (U/K)^{1/n}; \quad \theta = m/n \quad (2b)$$

where k_s and θ can be termed the steepness and concavity indices, respectively. Equation (2a) is often applied in analysis of channel profile data^{14,29–32}. Typical values of the concavity index (θ) for arguably equilibrium channel profiles in tectonically active fluvial landscapes given in Table 2 are consistent with theoretical predic-

Table 1 Hillslope gradient distributions in mountainous areas

Location	Mean slope (°)*	1 σ	Source
Altunshan, China	28.7	10.6	This work†
Altuntagh, China	2.8	9.3	This work†
Tien Shan, China	32	13.8	This work†
Mekong River, Yunnan Province, China	30.5	10.9	This work†
Nanga Parbat region, Pakistan	32	~10	Ref. 19
Dan Gabriel Mtns, California, USA	3.15	8.5	This work‡
Eastern White Mtns, California, USA	34	8.2	This work‡
Western White Mtns, California, USA	34	9.2	This work‡
King Range, California, USA	33	8.3	This work‡

* All measured slopes are averaged across 3 pixels and therefore represent minimum estimates.

† 90-m resolution digital elevation model (DEM) data (slopes strongly underestimated).

‡ 30-m resolution DEM data.

tions that the m/n ratio should fall in a narrow range near 0.5 ($0.35 \leq m/n \leq 0.6$) for fluvial erosion processes²⁶. Empirical values of θ outside the expected range¹⁴ probably reflect some combination of disequilibrium conditions, systematic downstream variation in K or U , or regression of data that cross the bedrock-alluvial transition. Under steady-state conditions and spatially invariant K and U , equation (1) can be integrated to find the equilibrium channel relief (R_f) of a bedrock river:

$$R_f = z(x_c) - z(L) = \left(\frac{U}{K} \right)^{\frac{1}{n}} k_a^{-\frac{m}{n}} \left(1 - \frac{hm}{n} \right)^{-1} \left(L^{1-\frac{hm}{n}} - x_c^{1-\frac{hm}{n}} \right); \quad \frac{hm}{n} \neq 1 \quad (3a)$$

$$R_f = z(x_c) - z(L) = \left(\frac{U}{K} \right)^{\frac{1}{n}} k_a^{-\frac{m}{n}} (\ln L - \ln x_c); \quad \frac{hm}{n} = 1 \quad (3b)$$

where L is the length of the bedrock stream measured from the divide, and x_c is the length of the hillslope plus any colluvial channel segment, also measured from the divide ($x_c \approx A_c^{1/2}$; see Table 2). Thus equation (3) gives the elevation drop on the bedrock channel only (R_f). We note that equation (3) shows that equilibrium bedrock channel relief varies inversely with K . Thus, for the case of a spatially constant increase in K , both tributary and trunk channel relief ultimately will be reduced in response to a shift to a more erosive climate (Fig. 2). This simple conclusion directly contradicts the commonly cited notion that increased precipitation (and presumably greater erosivity) causes greater relief along tectonically

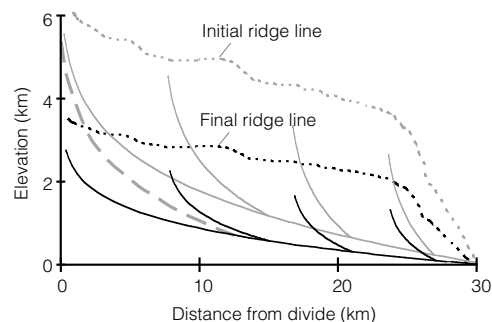


Figure 2 Equilibrium relief in fluvial landscapes. Solid grey lines show initial steady-state trunk and tributary channel profiles in a climate with low erosive potential (that is, low K in equation (1)). Dashed grey line shows the corresponding initial ridge-line topographic envelope. Solid black lines show final steady-state trunk and tributary channel profiles in a more erosive climate (high K). Dashed black line shows the corresponding final ridge-line topographic envelope. We have assumed that hillslope relief had already attained a maximum in the less erosive climate (rock uplift rate $\geq$ soil production rate²⁰). Comparison of these two steady-state conditions illustrates that a change to a more erosive climate must ultimately reduce relief in a fluvial landscape. Heavy long-dashed grey line illustrates the steepening of upper channel reaches (shown for trunk stream only) required to induce an increase in relief.

Table 2 Relief statistics in tectonically active fluvial landscapes

Field area	Crit. drainage area* A_c (10^3 m^2)	Avg. colluvial slope, S_c (m/m)	Fluvial relief, R_f/R_t (%)	Concavity index†, θ	No of drainages, N
King Range, California, USA (high uplift rate)	0.59 ± 0.20	0.54 ± 0.11	79 ± 7	0.40 ± 0.10	14
King Range, California, USA (low uplift rate)	0.72 ± 0.24	0.36 ± 0.05	80 ± 5	0.49 ± 0.10	7
Central Range, Taiwan	1.40 ± 0.48	0.63 ± 0.26	89 ± 6	0.41 ± 0.1	4

All uncertainties indicate 1σ -error bars.

* Defined by the break in slope–area scaling in longitudinal profile data.

† Defined by equation (2) in the text; reported values are fits to long profile data between A_c and the bedrock–alluvial transition only.

active mountain fronts^{12,13}. Despite the reduction in relief, the transition to a more erosive condition does increase rates of channel incision, and therefore rates of hillslope denudation by landsliding. However, the enhanced erosion is concentrated in the upper part of drainage basins as channels cut down to lower equilibrium gradients (Fig. 2). Simple geometry dictates that only a response to climate change that involves steepening of the upper reaches of the channel system in concert with accelerated incision of the trunk stream would in fact produce relief, as illustrated in Fig. 2.

Although the relief structure of landscapes is demonstrably altered during a transient response to a change in fluvial erosivity (Fig. 3), and also by non-uniform changes in K that may result from climate change¹⁴, we find that under most circumstances the steepening of headwater tributaries required to increase relief will not occur. Consider the transient response to a sudden, uniform change in K . Figure 3 shows that reduction of channel gradients first occurs along trunk streams and only later along tributary and headwater stream segments, resulting in a transient increase in channel concavity. However, as nowhere in the system are channel gradients increased, there is no accompanying increase in relief. Rather, the transient response is characterized by an upstream propagating wave of relief reduction (Fig. 3).

Spatially non-uniform increases in K generally produce similar effects. As long as K is everywhere increased, all channel gradients and therefore relief (both transient and steady state) will be reduced. If a given climate change causes K to be preferentially increased in downstream channel segments, transients will be shorter-lived and final steady-state channel profiles will resemble the transient profiles shown in Fig. 3, and relief will be similarly diminished. Conversely, if K is preferentially increased in upstream channel segments both channel gradient and concavity will be reduced, resulting in a greater relief reduction than in the uniform- K case. Thus, only a climate change which induces a decrease in K along headwater channel segments in concert with an increase in K downstream can be expected to increase relief. Although one may imagine a few

scenarios in which this could occur (for example, enhanced fluvial discharge but reduced frequency of debris flows in headwater colluvial channels), we are unaware of any evidence that such changes are representative of Quaternary climate change or that the resulting increase in relief would be significant. Further study of such non-uniform changes in erosivity is warranted.

Relief production in glaciated landscapes

The isolated spires, narrow ridges, long bare-rock hillslopes, and broad valleys of alpine terrain immediately give a qualitative impression that glaciers are highly efficient erosive agents and that glacial erosion produces considerable relief. This sense of the production of relief, now supported by data showing that erosion by fast-moving temperate glaciers can far outpace that expected for rivers³³, led Molnar and England⁵ to suggest that a transition to glacial erosion could induce significant isostatic peak uplift. However, such an effect has not been demonstrated empirically⁶, and even the premise that glacial erosion produces a net increase in relief has been challenged¹¹.

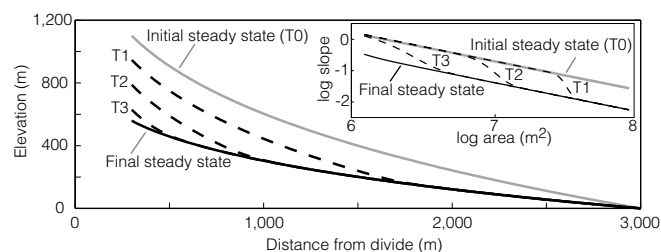


Figure 3 Numerical simulation of the transient response of a channel profile to a sudden, uniform increase in erosivity. (Shown is the effect of a two-fold increase in K in equation (1).) Inset shows the transient response in slope–area space ($\log$ – $\log$ scale). Both the main panel and inset shows the initial steady-state profile (T0; solid grey), three intermediate transient profiles, sequentially T1, T2, and T3 (black dashed lines), and the final steady-state profile (solid black). The transient response is characterized by an upstream-propagating wave of gradient, and therefore relief, reduction. Nowhere in the system are channel gradients increased (see inset). Note that if K were unchanged over the first kilometre of channel length and doubled only downstream of this point, the final steady-state solution would approximately coincide with the T2 transient profile.

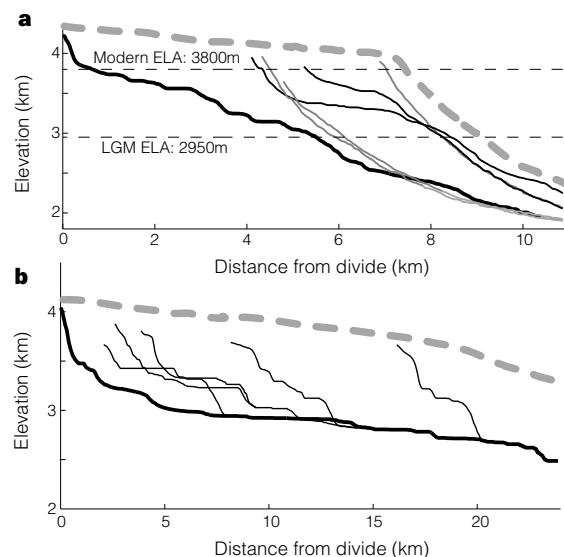


Figure 4 Relief structure of glaciated landscapes. **a**, Comparison of adjacent unglaciated (solid grey lines), partially glaciated (solid black lines), and fully glaciated (heavy black line) drainage basins along the eastern flank of the Sierra Nevada, near Lone Pine, California. Heavy dashed grey line illustrates the topographic envelope of ridges and peaks. Horizontal dashed lines indicate the modern and late Pleistocene positions of the equilibrium line altitude (ELA) in this part of the Sierra Nevada. Glacial valley lowering has been most effective at elevations between the modern and late Pleistocene ELAs, resulting in reduced valley profile concavity in the glaciated zone and a commensurate reduction of valley profile relief. Although there are some indications of ridge-to-valley relief increases at the mid-basin position, hillslope relief increases do not appear to have fully compensated for the reduction of valley profile relief. **b**, Relief structure of the Dinwoody Creek basin, northwestern Wind River Range, Wyoming, showing trunk valley profile (heavy black line), tributary profiles (black lines), and the topographic envelope of a pre-glacial low-relief surface⁶ (heavy dashed grey line). Hillslope relief on the trunk and tributary profiles averages ~ 400 m. Hanging valley relief ranges from 200 to 400 m.

As the mechanics of glacial erosion are less well known (there is no equivalent of a stream power erosion law for glaciers), our analysis of glaciated landscapes is perforce more qualitative and preliminary than the above argument for fluvial landscapes. Here we outline the characteristics of glacial erosion which are likely to alter the relief structure of a pre-existing fluvial landscape in an attempt to (1) place general constraints on the likely magnitude of relief production during a transition to glaciated conditions and (2) identify critical outstanding problems. We first consider relief-production and then relief-reduction mechanisms. Our analysis considers only temperate glaciers with wet bases, ignoring for the moment some potentially important effects of the possible occurrence of basal freezing along the upper reaches of thin, high-altitude alpine glaciers^{34,35}.

From study of alpine landforms (Fig. 4) in a range of geological and tectonic environments, and from consideration of the mechanics of alpine glacial erosion, we reason that glacial erosion may produce relief through four main mechanisms: valley widening, ice buttressing of rock slopes, and formation of hanging valleys and overdeepenings. Widening of V-shaped fluvial valleys is a natural consequence of erosion by thick masses of slowly moving ice³⁶. Narrow ridges between first-order side tributaries are erased as tributaries are replaced by hillslopes (Fig. 5a) and wide ridges between the main tributaries are reduced to narrow knife-edged ridges and spurs. The net effect is a decrease in drainage density with a commensurate lengthening of hillslopes, thus increasing the 'missing mass' in valleys. In addition, hillslope relief can be further increased during glacial incision (Fig. 5a) because of the stabilizing effect of buttressing by thick ice masses, despite long-term constraints on hillslope relief due to landsliding^{18,19}. Although landslides following glacial retreat are common, the long, oversteepened

bare-rock slopes of many formerly glaciated valleys (Fig. 4) indicate that this transient increase in hillslope relief is long lived compared to isostatic response times. The formation of hanging valleys effectively insulates ridge line lowering from lowering in the main valley (Fig. 5b), thus increasing tributary or 'valley' relief. However, the additional 'missing mass' associated with hanging-valley relief is limited to material removed along the axis of the trunk valley itself. Finally, glaciers are able to erode below fluvial base-level, as reflected in the formation of overdeepenings such as fjords, which adds an additional component of relief production.

Acting in direct opposition to these relief-production mechanisms, we identify three mechanisms by which glacial erosion may cause a reduction in some components of relief: concentration of erosion at higher elevations¹¹, reduction of fluvial erosion downstream of glaciers, and possible acceleration of summit-lowering rates in the near-glacial environment. Accelerated glacial erosion and the associated relief-generation mechanisms listed above naturally only occur above the ice limit, with most erosion focused around the position of the equilibrium line altitude (ELA), where ice flux reaches a maximum³⁷. As with climate-induced increases in fluvial erosion, glacial erosion is therefore concentrated in the upper parts of drainage basins and must contribute to a reduction of the elevation fall along the trunk valley profile (Fig. 5c). This effect is clearly shown in the longitudinal profiles of glacial valleys carved by wet-based temperate glaciers. Glacial valley profiles are not only significantly less steep but also less concave than fluvial bedrock channel profiles (Fig. 4a). Analyses of glaciated valley profiles in the western US reveal consistently lower concavity than typical fluvial valleys ($\theta = 0.18 \pm 0.08(1\sigma)$, determined for 16 glaciated drainages in the Sierra Nevada, Wind River, Bitterroot, and Wasatch Ranges, compared with 0.35–0.6 for bedrock fluvial channels). Also, aggradation caused by large increases in sediment flux associated with glacial erosion³³ will bury bedrock channels in alluvium; this will reduce rates of bedrock channel lowering below the glacial limit^{14,24}, thereby acting to further limit relief production along the trunk stream (Fig. 5c). Finally, periglacial processes may accelerate the rate of summit erosion and thus diminish any potential relief increases, although no data have as yet demonstrated this effect⁶.

Whether glacial erosion deepens valleys, and thus induces an isostatic uplift of peaks and ridges^{5,6}, or acts to plane off mountain ranges at (or somewhat above) the ELA¹¹ depends on the relative efficacy of the above relief-generation and relief-reduction mechanisms. Unfortunately, the mechanics of glacial erosion and relief production are insufficiently well known at present to permit definitive, quantitative answers to this question. But our preliminary observations in a wide range of glaciated landscapes lead us to suggest that relief increases attributable to most of the above-cited mechanisms scale with ice thickness, and therefore produce only limited additional relief. For instance, the potential increase of relief due to the ice buttressing effect and the width of parabolic glacial valley cross-sections both depend directly on ice thickness. Similarly, as the surfaces of tributary ice streams typically meet at the same elevation (that is, tributary ice falls are rare during full-glacial times), hanging valley relief must scale with the difference in ice thickness. If this ice-thickness hypothesis stands the test of time, and if no other as yet unrecognized relief-production mechanisms come to light, the implication is that average basin-wide relief production by wet-based temperate glaciers may be limited to several hundred metres (of the order of typical ice thicknesses).

Although the formation of hanging valleys and overdeepenings may entail the production of considerable amounts of vertical relief (with as many as 2–4 hanging valley or overdeepening steps between ridge lines and trunk stream valley bottoms in large basins), only a small percentage of the land surface areas is affected by this incision because most of the land surface is insulated behind hanging valley steps. The study by Small and Anderson⁶ of Quaternary relief production by glacial erosion in the Wind River Range, Wyoming,

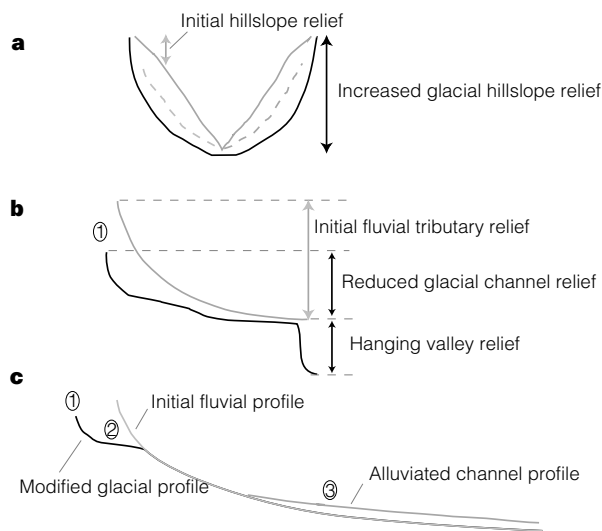


Figure 5 Glacial relief adjustment mechanisms. **a**, An initial V-shaped valley defined by ridge lines (solid grey) and channel profiles (dashed grey) of steep first-order tributaries, and a modified U-shaped glacial valley profile (solid black). Glacial valley widening and ice-buttressing of hillslopes together result in the obliteration of small tributary valleys, reduction in drainage density, and an increase in hillslope relief (indicated). **b**, An initial fluvial tributary channel profile and relief (grey lines) and modified glacial tributary valley and hanging valley profiles and relief (black lines). The formation of hanging valleys adds an important component of relief in glaciated landscapes, effectively insulating ridge lines from trunk valley incision. However, reductions in both slope and concavity of glacial valley profiles tends to reduce relief along tributary valley profiles. The net change in relief depends critically on the relative magnitude of hillslope relief increases above the tributary valleys (at location 1). **c**, Relief adjustments along longitudinal profiles of glaciated valleys are affected by: hillslope relief increases in the glaciated zone (location 1), reduced concavity and slope of valley profile in the glaciated zone (location 2), and alluviation of downstream fluvial bedrock channel reaches (location 3), which acts to further inhibit relief production.

serves as an excellent example of this effect.

Small and Anderson⁶ document glacial valley deepening of ~800 m into a pre-existing high-elevation but low-relief surface. Figure 4b shows the relief structure of a typical glacial valley in the Wind River Range. Most of the valley deepening is clearly accounted for by enhanced hillslope relief (~400 m) and the drop on hanging valleys (200–400 m). Despite the great depth of the trunk glacial valleys, however, the formation of hanging valleys restricts the total volume of mass removal associated with this incision. Accordingly, by dividing their estimate of the total volume of 'missing mass' by land surface area, Small and Anderson⁶ compute an average Quaternary relief production of the order of 300 m. The Late Pleistocene glaciers that carved these valleys were relatively thick, of the order of 300–400 m, as indicated by erosional trimlines and estimations based on valley slope³⁴; this is consistent with the observed hillslope and hanging valley relief and the hypothesis put forward above. Moreover, for the specific case of the Wind River Range, Small and Anderson conclude that there has been negligible isostatic uplift of summit surfaces during the Quaternary—despite the nearly 300 m of relief production—due to the action of slow but non-negligible (~0.01 mm yr⁻¹) summit lowering³⁸ and flexural distribution of the isostatic response to the erosional unloading.

Anticipated effect of climate change on topographic relief

In tectonically active mountain belts, the increase in relief due to accelerated erosion induced by climate change, and therefore the commensurate flexural-isostatic uplift of peaks and ridges, is likely to be minimal. Indeed, in almost all non-glacial landscapes an increase in the erosivity of the fluvial system is anticipated to lead to a reduction in both trunk stream and tributary relief. When coupled with the constraint that hillslope relief rapidly attains a maximal condition in active orogens, this observation implies that ridge to valley bottom relief will actually decrease under these conditions. We note that relief increase is possible if (and only if) a given climate change induces a decrease in erosivity along headwater channel segments in concert with a simultaneous increase in erosivity farther downstream. Aspects of relief evolution in non-glacial environments that merit further study are: (1) possible increases in the relief on poorly understood colluvial, debris-flow channel segments²⁴, (2) the magnitude and duration of possible relief increases due to transient increases in channel concavity associated with non-uniform changes in the coefficient of erosion, K , and (3) the quantitative relationship between climate variables and K .

The onset of glacial erosion is likely to involve competing effects of relief production over short wavelengths (increases in hillslope relief, valley widening, and formation of hanging valleys and overdeepenings) and relief reduction over long wavelengths (reduction of relief on trunk and tributary valley profiles). Further study is needed to quantify fully the net effects of the various relief generation and reduction mechanisms outlined here. In particular, the upper limit on transient glacial hillslope relief increases is an important unknown. In addition, if the upper reaches of thin, high-altitude alpine glaciers become frozen to their beds, ridges and peaks may be protected from erosion and glacial valley long profiles may become more concave³⁵, adding a potentially important component to overall glacial relief production. But for the case of wet-based temperate glaciers, we argue that relief generation associated with each of the various glacial relief production mechanisms scales with ice thickness. Although the formation of a cascade of hanging valleys may entail the production of considerable amounts of vertical relief, the added 'missing mass' attributable to this incision is restricted by the limited areal extent of the trunk valleys themselves. If correct, this argument implies that net production of relief in these systems is probably limited to several hundred metres. It thus seems unlikely that significant amounts of isostatic peak uplift would be induced by a climate change that involves either enhanced fluvial erosivity or a transition to wet-based glacial erosion. □

Received 17 March; accepted 5 July 1999.

- Koons, P. O. Modeling the topographic evolution of collisional belts. *Annu. Rev. Earth Planet. Sci.* **23**, 375–408 (1995).
- Beaumont, C., Kamp, P., Hamilton, J. & Fullsack, P. The continental collision zone, South Island, New Zealand: Comparison of geodynamical models and observations. *J. Geophys. Res.* **101**, 3333–3359 (1996).
- Gilchrist, A. R., Summerfield, M. A. & Cockburn, H. A. P. Landscape dissection, isostatic uplift, and the morphologic development of orogens. *Geology* **22**, 963–966 (1994).
- Montgomery, D. R. Valley incision and the uplift of mountain peaks. *J. Geophys. Res.* **99**, 13913–13921 (1994).
- Molnar, P. & England, P. Late Cenozoic uplift of mountain ranges and global climate change: Chicken or egg? *Nature* **346**, 29–34 (1990).
- Small, E. E. & Anderson, R. S. Pleistocene relief production in Laramide mountain ranges, western United States. *Geology* **26**, 123–126 (1998).
- Stuwe, K., White, L. & Brown, R. The influence of eroding topography on steady-state isotherms. Application to fission track analysis. *Earth Planet. Sci. Lett.* **124**, 63–74 (1994).
- Mancktelow, N. S. & Grasemann, B. Time-dependent effects of heat advection and topography on cooling histories during erosion. *Tectonophysics* **270**, 167–195 (1997).
- Raymo, M. E. & Ruddiman, W. F. Tectonic forcing of late Cenozoic climate. *Nature* **359**, 117–122 (1992).
- Small, E. E. & Anderson, R. S. Geomorphically driven Late Cenozoic rock uplift in the Sierra Nevada, California. *Science* **270**, 277–280 (1995).
- Brozovic, N., Burbank, D. & Meigs, A. Climatic limits on landscape development in the Northwestern Himalaya. *Science* **276**, 571–574 (1997).
- Fielding, E. J., Isacks, B. L., Barazangi, M. & Duncan, C. How flat is Tibet? *Geology* **22**, 163–167 (1994).
- Masek, J. G., Isacks, B. L., Gubbels, T. L. & Fielding, E. J. Erosion and tectonics at the margins of continental plateaus. *J. Geophys. Res.* **99**, 13941–13956 (1994).
- Sklar, L. & Dietrich, W. E. in *Rivers Over Rock: Fluvial Processes in Bedrock Channels* (eds Tinkler, K. J. & Wohl, E. E.) 237–260 (AGU Press, Washington DC, 1998).
- Montgomery, D. R. et al. Distribution of bedrock and alluvial channels in forested mountain drainage basins. *Nature* **381**, 587–589 (1996).
- Montgomery, D. R. & Buffington, J. M. Channel-reach morphology in mountain drainage basins. *Geol. Soc. Am. Bull.* **109**, 596–611 (1997).
- Montgomery, D. R. & Dietrich, W. E. Channel initiation and the problem of landscape scale. *Science* **255**, 826–830 (1992).
- Schmidt, K. M. & Montgomery, D. R. Limits to relief. *Science* **270**, 617–620 (1995).
- Burbank, D. W. et al. Bedrock incision, rock uplift and threshold hillslopes in the northwestern Himalayas. *Nature* **379**, 505–510 (1996).
- Heimsath, A. M., Dietrich, W. E., Nishizumi, K. & Finkel, R. C. The soil production function and landscape equilibrium. *Nature* **388**, 358–361 (1997).
- Hovius, N., Stark, C. P. & Allen, P. A. Sediment flux from a mountain belt derived by landslide mapping. *Geology* **25**, 231–234 (1997).
- Hovius, N., Stark, C. P., Chu, H. T. & Lin, J. C. Supply and removal of sediment in a landslide-dominated mountain belt: Central Rang, Taiwan. *J. Geol.* (in the press).
- Tucker, G. E. & Bras, R. L. Hillslope processes, drainage density, and landscape morphology. *Wat. Resour. Res.* **34**, 2751–2764 (1998).
- Howard, A. D. in *Rivers Over Rock: Fluvial Processes in Bedrock Channels* (eds Tinkler, K. & Wohl, E. E.) 297–319 (AGU Press, Washington DC, 1998).
- Rivers Over Rock: Fluvial Processes in Bedrock Channels* (eds Tinkler, K. & Wohl, E. E.) 323 (AGU Press, Washington DC, 1998).
- Whipple, K. X. & Tucker, G. E. Dynamics of bedrock channels in active orogens: Implications for height limits of mountain ranges, landscape response timescales, and research needs. *J. Geophys. Res.* **104**, 17661–17674 (1999).
- Hack, J. T. Studies of longitudinal stream profiles in Virginia and Maryland. *US Geol. Surv. Prof. Pap.* **294-B**, 97 (1957).
- Howard, A. D., Seidl, M. A. & Dietrich, W. E. Modeling fluvial erosion on regional to continental scales. *J. Geophys. Res.* **99**, 13971–13986 (1994).
- Tarboton, D. G., Bras, R. L. & Rodriguez-Iturbe, I. Scaling and elevation in river networks. *Wat. Resour. Res.* **25**, 2037–2051 (1989).
- Montgomery, D. R. & Foufoula-Georgiou, E. Channel network representation using digital elevation models. *Wat. Resour. Res.* **29**, 1178–1191 (1993).
- Willgoose, G. A physical explanation for an observed area-slope-elevation relationship for catchments with declining relief. *Wat. Resour. Res.* **30**, 151–159 (1994).
- Snyder, N. P., Whipple, K. X., Tucker, G. E. & Merritts, D. J. Landscape response to tectonic forcing: DEM analysis of stream profiles in the Mendocino triple junction region, northern California. *Geol. Soc. Am. Bull.* (in the press).
- Hallet, B., Hunter, L. & Bogen, J. Rates of erosion and sediment evacuation by glaciers: a review of field data and their implications. *Glob. Planet. Change* **12**, 213–235 (1996).
- Paterson, W. S. B. *The Physics of Glaciers* (Pergamon, New York, 1994).
- Oerlemans, J. Numerical experiments on large-scale glacial erosion. *Z. Gletscherskunde Glazialgeol.* **20**, 107–126 (1984).
- Harbor, J. M. Numerical modeling of the development of U-shaped valleys by glacial erosion. *Geol. Soc. Am. Bull.* **104**, 1364–1375 (1992).
- Andrews, J. T. Glacier power, mass balances, velocities and erosion potential. *Z. Geomorphol. N. F. Suppl. Bd.* **13**, 1–17 (1972).
- Small, E. E., Anderson, R. S., Repka, J. L. & Finkel, R. Erosion rates of alpine bedrock summit surfaces deduced from in situ ¹⁰Be and ²⁶Al. *Earth Planet. Sci. Lett.* **150**, 413–425 (1997).

Acknowledgements

This work was supported in part by the NSF (Continental Dynamics) and NASA (K.X.W.). We thank G. Tucker for discussions of relief evolution in active orogens, and D. Burbank, P. Molnar and E. Small for comments and suggestions.

Correspondence and requests for materials should be addressed to K.X.W. (e-mail: kxw@mit.edu).

Continuous heating of a giant X-ray flare on Algol

J. H. M. M. Schmitt* & F. Favata†

* Universität Hamburg, Hamburger Sternwarte, Gojenbergsweg 112, D-21029 Hamburg, Germany

† European Space Agency-European Space Research and Technology Centre (ESA-ESTEC), Space Science Department, ESTEC, Postbus 299, NL-2200 AG Noordwijk, The Netherlands

Giant stellar flares can release large amounts of energy within a few days^{1–7}: X-ray emission alone can be up to ten per cent of the star's bolometric luminosity. These flares exceed the luminosities of the largest solar flares by many orders of magnitude, which suggests that the underlying physical mechanisms supplying the energy are different from those on the Sun. Magnetic coupling between the components in a binary system or between a young star and an accretion disk has been proposed^{3,7–9} as a prerequisite for giant flares. Here we report X-ray observations of a giant flare on Algol B, a giant star in an eclipsing binary system. We observed a total X-ray eclipse of the flare, which demonstrates that the plasma was confined to Algol B, and reached a maximum height of 0.6 stellar radii above its surface. The flare occurred around the south pole of Algol B, and energy must have been released continuously throughout its life. We conclude that a specific extrastellar environment is not required for the presence of a flare, and that the processes at work are therefore similar to those on the Sun.

X-ray and radio observations have produced copious evidence for magnetic-field-related activity on nearly all cool stars with outer surface convection zones^{10,11}. Magnetic activity on other stars is usually interpreted by analogy with the Sun, viewing the observed stellar emissions as coming from scaled-up versions of directly observable solar features. The physical validity of this approach is debatable, particularly for the observed extremes of stellar activity, which exceed the corresponding solar emissions by at least five orders of magnitude.

Magnetic reconnection is important in the theory of solar flares¹², yet the application of the theory to solar observations produces ambiguous evidence for the occurrence of reconnection^{13,14}. Reconnection is a local phenomenon; the difficulty lies in explaining how energy can be released swiftly and coherently from rather small volumes. This becomes even more difficult when discussing the energetics of giant stellar flares. At peak, such flares can release up to a few per cent of the quiescent bolometric luminosity, and their overall energy release equals the cumulative output of a few hours of quiescent luminosity; there must be, therefore, an efficient way to store, and then release, vast amounts of energy. Extreme levels of activity are observed in close binary systems and in very young stars, where the magnetic field topology is likely to differ from that of the Sun. In a close binary system, magnetic field lines may connect the two components, enabling plasma to be magnetically confined over a much larger interbinary volume; similarly, in a star surrounded by an accretion disk, magnetic field lines originating in the photosphere may thread the disk and magnetic stresses may be built up by shearing motions. It has therefore been proposed that the specific environment of such systems is a prerequisite for the occurrence of giant flares^{3,7–9}.

In eclipsing binaries, size information can be obtained from light curves. The eclipsing binary Algol is one of the nearest and brightest X-ray eclipsing systems; the relevant system parameters¹⁵ are given in Table 1. Frequent X-ray flares have been reported on Algol^{14,16,17}, and the absence of an X-ray eclipse at optical secondary minimum

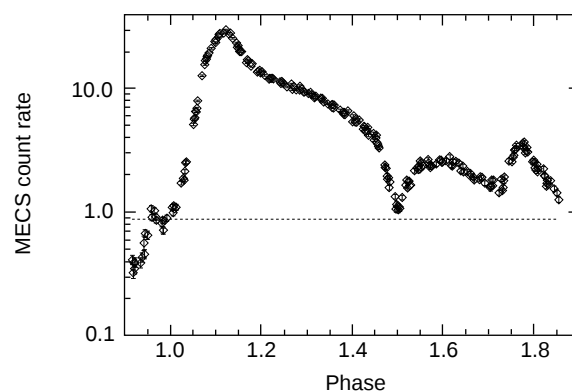


Figure 1 X-ray light curve of Algol measured with the BeppoSAX MECS2 and MECS3 detectors. Measurements were made in the 1.6–10 keV band between 30 August 1997 3:04 UT and 1 September 1997 20:32 UT with phase ϕ calculated from the ephemeris $JD = 2445739.003 + 2.8673285 \times E$ (where E is an integer) for the times of primary minimum¹⁵; a hundredth of a phase corresponds to 41.3 minutes. The phasing is such that for $\phi_E = E + 0.5$, the X-ray-dark early-type primary is in front of the X-ray-emitting late-type secondary. Note the huge flare starting at phase $\phi \sim 1.0$ with a rise time of about 8.3 hours and peaking at $\phi \sim 1.12$. A rapid initial decay until $\phi \sim 1.25$ is followed by an exponential decay. A clear eclipse of the flaring plasma is seen at $\phi \sim 1.5$, when the early-type primary is in front of the late-type secondary. The dotted line indicates an estimate of the quiescent out-of-flare rate, extrapolated from the observed Rosat all-sky survey count rate.

has been interpreted as evidence for a corona with a scale height of more than a stellar radius¹⁸. A 2.9-day X-ray observation of Algol, covering the whole binary orbit, was carried out with the Italian BeppoSAX satellite¹⁹. Here we discuss only the medium energy concentrator spectrometer (MECS) light curve in the energy band 1.6–10 keV. In Fig. 1 we show the phase-folded BeppoSAX MECS light curve of Algol, which is dominated by a huge flare lasting almost through the entire observation. The observed peak luminosity is $3 \times 10^{32} \text{ erg s}^{-1}$ (at least 1.2% of the late-type component's quiescent luminosity), and integration over all of the available BeppoSAX light curves yields a total observed X-ray energy release $E_{X\text{-ray,tot}} \approx 1.5 \times 10^{37} \text{ erg}$ in the 0.1–10 keV band. The flare is thus at least as energetic as the largest flares observed on RS CVn systems, T Tauri and proto-stars^{1–7}.

At phase $\phi = 1.5$ a count rate minimum is observed (Fig. 1). The only plausible interpretation of this light curve feature is an X-ray eclipse of the flaring plasma by the early-type primary. Residual X-ray flux persists at minimum, but a number of indications suggest that the flux comes from Algol's quiescent (unocculted) corona, and not from the flare: at $\phi = 1.5$ the light curve is straight while it decreases and increases before and after that phase, at minimum the observed count rate is close to the count rates before flare onset, and also the observed count rate at minimum is very similar to the mean Rosat all-sky survey count rate. It thus seems reasonable to assume that the eclipse of the X-ray flare is total. In order to validate this assumption, we attempted to determine the appropriate quiescent emission level at phase $\phi = 1.5$. All values between 0.4 and 1.5 counts s^{-1} can be chosen and need to be compared to the observed minimum rate of 1.08 counts s^{-1} (Fig. 2a). We thus estimate that at least 90% of the flare emission was occulted. As a temperature increase was detected only when the count rate was above 1.1 counts s^{-1} , we assume henceforth that the eclipse of the flare was total.

Three important features of the BeppoSAX X-ray light curve are evident: (1) a total eclipse of the X-ray flare caused by the X-ray dark early-type primary; (2) a symmetrical pattern of eclipse ingress and egress with a shallow light curve decrease of $\sim 20\%$ followed by a steep decrease of the remaining 80% on ingress, reversed on egress;

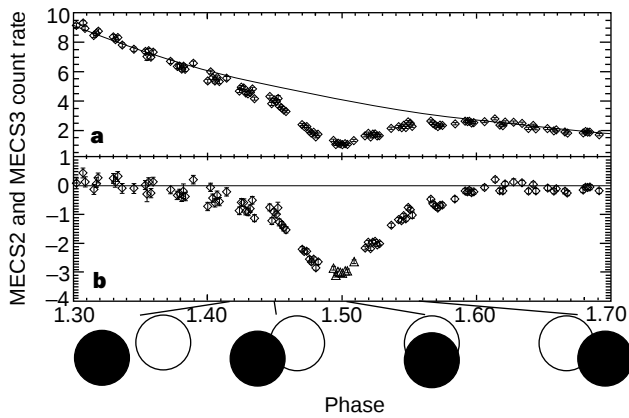


Figure 2 The MECS count rate versus phase in the interval 1.3–1.7. **a**, Count rate versus phase; the solid line represents an exponential fit to the pre- and post-eclipse light curve. **b**, Count rate versus phase in the interval 1.3–1.7 with exponential decay (shown in **a**) removed; the zero line is shown. The flare eclipse starts at $\phi \sim 1.39$ with a somewhat shallow decay; at $\phi \sim 1.451$ a sharp decay starts (first contact). Totality begins at $\phi \sim 1.488$ (second contact) and ends at $\phi \sim 1.505$ (third contact). The light curve increases quickly until $\phi \sim 1.545$ (fourth contact) and then more slowly until $\phi \sim 1.605$, when the flare eclipse is over. Below **b** the viewing geometry is shown at phases $\phi \sim 1.42$, 1.45, 1.50 and 1.55; the filled circle represents Algol A as it moves across Algol B.

(3) no evidence for any occultation of the flare plasma by the late-type secondary (no rotational modulation; note that because of synchronous rotation two different stellar hemispheres are in view at $\phi \approx 1.5$ during minimum and at $\phi \sim 1.0$ when the flare erupts). The occurrence of an eclipse at $\phi = 1.5$ indicates that the flare is associated with the secondary; the absence of rotational modulation then implies that the flare plasma must be located near one of the poles. The eclipse totality restricts the maximum height above the surface to $d_{AB} \sin(i) = R_*$, where d_{AB} is the distance between the binary components Algol A and Algol B, i is the angle between orbit plane normal and line of sight, and R_* is the radius of the star (Table 1), considerably less than a stellar radius. A three-dimensional view of the admissible flare locations is displayed in Fig. 3. The flare undoubtedly occurred above the south pole of Algol B and not in the interbinary region.

For the maximum available volume in Fig. 3 we derive a value of $V_{\max} = 1.5 \times 10^{33} \text{ cm}^3$; the actual flare is likely to occupy only some fraction $f_{\max}$ of $V_{\max}$. From spectral analysis²⁰ the total emission measure at the flare peak, $EM = n^2 V = 1.33 \times 10^{55} \text{ cm}^{-3}$, and the temperature $T = 1 \times 10^8 \text{ K}$ can be determined. Hence, the flare plasma density $n = 9.4 \times 10^{10} / f_{\max}^{1/2} \text{ cm}^{-3}$ and the thermal pressure is $2.6 \times 10^3 / f_{\max}^{1/2} \text{ dyn cm}^{-2}$. Given these values the radiative cooling time¹⁷ $\tau_{\text{rad}} = 3kT/nP(T)$, where k is Boltzmann's constant, and the cooling function $P(T) = 10^{-24.73} T^{1/4}$ (in CGS units), becomes $\tau_{\text{rad}} = 23,600 \times f_{\max}^{1/2}$ seconds. This is much smaller than the observed light curve decay time scale $\tau_{\text{dec}} \approx 60,000$ seconds, regardless of $f_{\max}$. It seems logical to conclude that continued heating through continued energy release must be present throughout the flare. Of course, the flare plasma may cool conductively. Determining the conductive cooling time scale¹⁷, we find

Table 1 System parameters for Algol (β Per)

Parameter	Primary (A)	Secondary (B)	System
Mass (g)	6.54×10^{33}	1.64×10^{33}	
Radius (cm)	2.15×10^{11}	2.29×10^{11}	
Spectral type	B8V	K2 III	
Luminosity (erg s^{-1})	5.95×10^{35}	2.44×10^{34}	
Rotation period (days)	2.8673	2.8673	
Orbital period (days)			2.8673
Distance (cm)			1.02×10^{12}

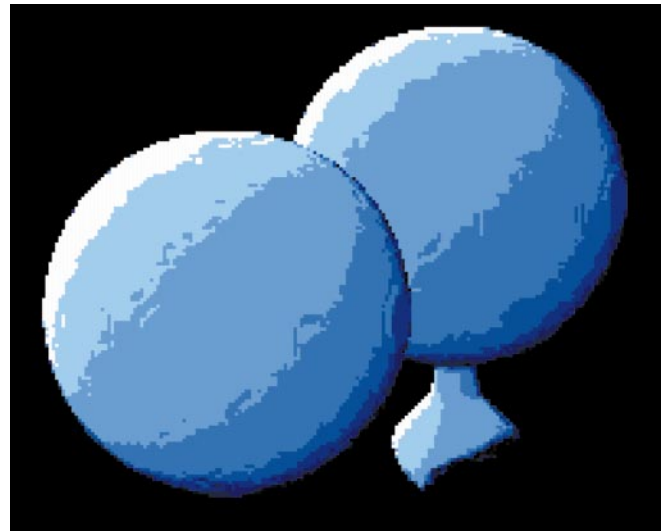


Figure 3 Three-dimensional visualisation of the Algol system including the maximum volume containing the flare. The two binary components Algol A (foreground) and Algol B (background) are shown as seen from our line of sight at phase $\phi = 1.45$. The extension below the south pole of Algol B is the computed locus of all points satisfying the following requirements: (1) they must be totally eclipsed by Algol A between second and third contact; (2) they must not be eclipsed by Algol A and not occulted by Algol B itself before first and after fourth contact; (3) they must be visible either between first and second or between third and fourth contact but not both; (4) assuming that the flaring plasma originates from the surface of Algol B, along any radial line of sight from the surface of Algol B to the point under consideration there should be no more than twenty per cent rotational modulation. These requirements define the hose-like structure above the south pole of Algol B, within which the flare must have occurred. The flare is clearly associated with Algol B and did not occur in the interbinary region.

$\tau_{\text{cond}} = 2,530 \times f_{\max}^{-1/2} \text{ sec}$ by setting the adopted length scale $L = V_{\max}^{1/3}$. This value of τ_{cond} is smaller than τ_{rad} for large filling factors; however, τ_{cond} is sensitively dependent on L , and we are thus unable to prove that $\tau_{\text{cond}} < \tau_{\text{rad}}$ holds.

Small heights above the surface and continued heating are characteristic properties of 'two-ribbon flares' on the Sun²¹. The energy for continued heating is derived from the reconnection of (non-potential) magnetic fields of opposing polarity. If all of the released energy E_{released} is derived from the magnetic energy in the volume $V_{\max}$, the minimum required magnetic field strength $B_{\min}$ can be computed by equating $B_{\min}^2 V_{\max} / 8\pi = E_{\text{released}} f_{X-\text{ray}}$, where $f_{X-\text{ray}}$ denotes the observed fraction of the released energy. We find $B = 500 / (f_{X-\text{ray}})^{1/2} \text{ G}$ to be the (non-potential) pre-reconnection magnetic field that needs to be annihilated to meet the energy requirements. If, in addition, the hot X-ray emitting plasma generated in the reconnection process must be magnetically confined, field strengths of at least $B = 256 / f_{\max}^{1/4} \text{ G}$ after reconnection are required.

Our finding, that the Algol flare occurred above the south pole of Algol B, corresponds well to the presence of 'polar spots' in many rapidly rotating active binary systems^{22,23}. While no magnetic field measurements or Doppler images have been made of Algol B, it is reasonable to assume the presence of such polar spots on Algol B. Photospheric magnetic field strengths²⁴ on late-type stars are typically in the range 1000–2000 G, and these values should also apply to Algol B. These values are clearly a strict upper limit to the coronal magnetic field strengths.

A solar-like reconnection scenario thus appears to be the most natural explanation of the observations of the giant flare on Algol. For physical consistency, magnetic fields of 500 G–1000 G, much larger than in the corona of the Sun²¹, must be present in the corona of Algol B at heights of up to half a stellar radius, extending over a volume of $> 10^{33} \text{ cm}^3$ and reconnected in the course of the flare. The

flaring plasma is associated only with the late-type star, and the binary nature of Algol is irrelevant, at least for this giant flare. The flare does, however, also display several features not observed in solar flares. Solar flares are never observed in polar regions; instead they are confined to the active region belt at lower heliographic latitudes. The thermal plasma in the giant flare on Algol is dominated by temperatures of 1×10^8 K; such temperatures are occasionally observed in solar flares^{25,26} as 'super hot plasmas', but they contain only small fractions of the overall emission measure. The derived minimum plasma density is similar to the plasma densities of many solar flares²⁷, which would require a volume close to $V_{\max}$ to satisfy the observed energy budget. Higher densities with correspondingly smaller volumes can, of course, not be excluded; however, much higher densities quickly lead to implausibly large thermal pressures. Thus, the real challenge to theory is the construction of physically consistent reconnection-based flare models with the observed stellar parameters and clarify the effect of giant flares on the mass and angular momentum loss of active stars. $\square$

Received 17 February; accepted 19 July 1999.

- Kürster, M. & Schmitt, J. H. M. Forty days in the life of CF Tucanae (=HD 5303). The longest stellar X-ray flare observed with ROSAT. *Astron. Astrophys.* **311**, 211–229 (1996).
- Foing, B. H. et al. in *IAU Colloq. 153: Magnetodynamic Phenomena in the Solar Atmosphere—Prototypes of Stellar Magnetic Activity* (eds Uchida, Y., Kosugi, T. & Hudson, H. S.) 283–284 (Kluwer Academic, Dordrecht, 1996).
- Graffagnino, V. G., Wonnacott, D. & Schaeidt, S. HR 5110 superflare: an interbinary flare identified? *Mon. Not. R. Astron. Soc.* **275**, 129–142 (1995).
- Ottmann, R. & Schmitt, J. H. M. ROSAT observation of a giant X-ray flare on Algol: evidence for abundance variations? *Astron. Astrophys.* **307**, 813–823 (1996).
- Preibish, T., Zinnecker, H. & Schmitt, J. H. M. ROSAT-detection of a giant X-ray flare on LkH- α 92. *Astron. Astrophys.* **279**, L33–L36 (1993).
- Preibish, T., Neuhauser, R. & Alcalá, J. M. A giant X-ray flare on the young star P1724. *Astron. Astrophys.* **304**, L13–L16 (1996).
- Grosso, N., Montmerle, T., Feigelson, E. D., Andre, P., Casanova, S. & Gregorio-Hetem, J. An X-ray superflare on an infrared protostar. *Nature* **387**, 56–58 (1997).
- Ferreira, J. M. On the problem of very energetic flares in binary systems. *Astron. Astrophys.* **335**, 248–254 (1998).
- Van den Oord, G. H. J. Filament support and flares in binaries. *Astron. Astrophys.* **205**, 167–180 (1988).
- Schmitt, J. H. M., Fleming, T. A. & Giampapa, M. S. The X-ray view of the low-mass stars in the solar neighborhood. *Astrophys. J.* **450**, 392–400 (1995).
- Schmitt, J. H. M. Coronae on solar-like stars. *Astron. Astrophys.* **318**, 215–230 (1997).
- Forbes, T. G. in *Magnetic Reconnection in the Solar Atmosphere* (eds Bentley, R. D. & Mariska, J. T.) Vol. 111, 259–267 (ASP Conf. Ser., 1997).
- Hudson, H. S. & Khan, J. I. in *Magnetic Reconnection in the Solar Atmosphere* (eds Bentley, R. D. & Mariska, J. T.) Vol. 111, 135–144 (ASP Conf. Ser., 1997).
- Tsuneta, S. in *Magnetic Reconnection in the Solar Atmosphere* (eds Bentley, R. D. & Mariska, J. T.) Vol. 111, 409–418 (ASP Conf. Ser., 1997).
- Al-Naimiy, H. M. K., Mutter, A. A. & Flaih, H. A. UVB photometry and light-curve analysis of Algol. *Astrophys. Space Sci.* **108**, 227–236 (1985).
- Stern, R. A., Uchida, Y., Tsuneta, S. & Nagase, F. Ginga observations of X-ray flares on Algol. *Astrophys. J.* **400**, 321–329 (1992).
- Van den Oord, G. H. J. & Mewe, R. The X-ray flare and the quiescent emission from Algol as detected by XOSAT. *Astron. Astrophys.* **213**, 245–260 (1989).
- White, N. E., Culhane, J. L., Parmar, A. N., Kellett, B. J., Kahn, S., Van den Oord, G. H. J. & Kuijpers, J. An XOSAT observation of quiescent and flare coronal X-ray emission from Algol. *Astrophys. J.* **301**, 262–274 (1997).
- Boella, G., Butler, R. C., Perola, G. C., Piro, L., Scarsi, L. & Bleeker, J. A. M. The medium-energy concentrator spectrometer on board the BeppoSAX X-ray astronomy satellite. *Astron. Astrophys. (suppl.)* **122**, 327–340 (1997).
- Favata, F. & Schmitt, J. H. M. Spectroscopic analysis of a superhot giant flare observed on Algol by BeppoSAX. *Astron. Astrophys.* (accepted).
- Tsuneta, S. Structure and dynamics of magnetic reconnection in a solar flare. *Astrophys. J.* **456**, 840–849 (1996).
- Strassmeier, K. G. in *176th IAU Symp. on Stellar Surface Structure* (eds Strassmeier, K. G. & Linsky, J. L.) 289–298 (Kluwer Academic, Dordrecht, 1996).
- Byrne, P. B. in *176th IAU Symp. on Stellar Surface Structure* (eds Strassmeier, K. G. & Linsky, J. L.) 299–303 (Kluwer Academic, Dordrecht, 1996).
- Saar, S. H. in *IAU Colloq. 153: Magnetodynamic Phenomena in the Solar Atmosphere—Prototypes of Stellar Magnetic Activity* (eds Uchida, Y., Kosugi, T. & Hudson, H. S.) 367–374 (Kluwer Academic, Dordrecht, 1996).
- Masuda, S., Kosugi, S., Hara, H., Tsuneta, S. & Ogawara, Y. A loop-top hard X-ray source in a compact solar flare as evidence for magnetic reconnection. *Nature* **371**, 495–497 (1994).
- Hudson, H. S. & Nitta, N. in *High Energy Solar Physics* (eds Ramaty, R., Mandzhavidze, N. & Hua, X.-M.) **374**, 285–293 (AIP Conf. Proc., 1996).
- Bray, R. J., Cram, L. E., Durrant, C. J. & Loughhead, R. E. Plasma loops in the solar corona. *Cambridge Astrophys. Ser.* **18**, 239 (1991).

Correspondence and requests for materials should be addressed to J.H.M.M.S. (e-mail: jschmitt@hs.uni-hamburg.de).

Conical dislocations in crumpling

Enrique Cerda*, Sahraoui Chaleb†, Francisco Melo* & L. Mahadevan†

* Departamento de Física de la Universidad de Santiago de Chile, Avda Ecuador 3493, Casilla 307 Correo 2, Santiago, Chile

† Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

A crumpled piece of paper is made up of cylindrically curved or nearly planar regions folded along line-like ridges, which themselves pivot about point-like peaks; most of the deformation and energy is focused into these localized objects. Localization of deformation in thin sheets is a diverse phenomenon^{1–6}, and is a consequence of the fact⁷ that bending a thin sheet is energetically more favourable than stretching it. Previous studies^{8–11} considered the weakly nonlinear response of peaks and ridges to deformation. Here we report a quantitative description of the shape, response and stability of conical dislocations, the simplest type of topological crumpling deformation. The dislocation consists of a stretched core, in which some of the energy resides, and a peripheral region dominated by bending. We derive scaling laws for the size of the core, characterize the geometry of the dislocation away from the core, and analyse the interaction between two conical dislocations in a simple geometry. Our results show that the initial stages of crumpling (characterized by the large deformation of a few folds) are dominated by bending only. By considering the response of a transversely forced conical dislocation, we show that it is dynamically unstable above a critical load threshold. A similar instability is found for the case of two interacting dislocations, suggesting that a cascade of related instabilities is responsible for the focusing of energy to progressively smaller scales during crumpling.

To probe volume-restricting deformations that lead to localization in a simple setting, we consider a circular sheet of diameter $2R_p$ that is forced axially by a distance d into a rigid cylindrical hoop of diameter $2R < 2R_p$, in a manner similar to pushing coffee-filter paper into a funnel (Fig. 1). Owing to the energetically prohibitive

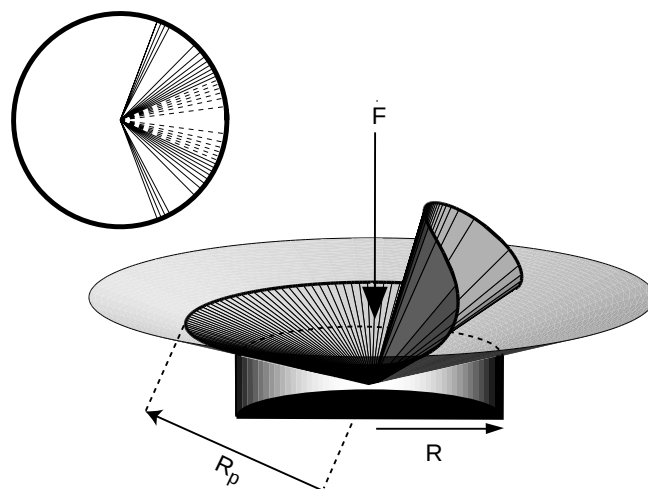


Figure 1 Geometry of an ideal conical dislocation. The shape is computed by solving the Euler–Lagrange equations corresponding to minimizing $L(\psi)$ in equation (1), with $\epsilon = 0.365$. The figure also shows the parameters that are experimentally controllable: R , the radius of the support; R_p , the radius of the sheet (depicted as a ruled surface); and the transverse applied force, F . At top left we show the generators of the conical surface: the solid lines correspond to the negatively curved region (concave-up) and the dotted lines correspond to the positively curved region.

requirement of making an axisymmetric cone, the sheet deforms into a non-axisymmetric conical surface that is in partial contact with the hoop^{10,11}. This surface has a single plane of symmetry and is isometric to the plane everywhere except near its tip or core; that is, it is a near-perfect developable cone^{5,12}. Only in the vicinity of the tip is the sheet stretched appreciably; elsewhere it is bent without stretching. This surface is also a rotational edge-dislocation, or conical dislocation, according to the Volterra classification^{11,13} and affords the simplest example of strain localization.

As this conical surface is nearly developable, its geometry can be characterized in terms of its generators¹⁴ away from the tip. These generators are most easily viewed using light reflected from the surface of a painted sheet that is illuminated with linearly polarized light incident along the axis of deformation. When viewed through a cross-polarizer, this leads to images such as that shown in Fig. 2. Away from the tip, the isochromatic lines (along which there is no radial curvature) are straight and form the generators of the perfect developable cone¹¹; close to the tip, they have a well defined radius of curvature R_c , that characterizes the size of the core where there is appreciable stretching. We determine R_c by fitting a parabola to the curved isochromatic line passing through the core. In Fig. 3a, we plot R_c as a function of the deformation characterized by $\epsilon = d/R$. The data are best fitted by two power laws; for $\epsilon \leq 0.1$, $R_c \propto \epsilon^{-0.33 \pm 0.01}$, while for $\epsilon \geq 0.1$, $R_c \propto \epsilon^{-0.50 \pm 0.02}$.

To understand these results, we consider the deformation of a circular sheet into a conical surface using eulerian coordinates. We let (ρ, θ) be the polar coordinates of a point in the horizontal plane defined by the edge of the supporting hoop, so that the vertical displacement of the sheet is $\rho\psi(\theta)$ where $\psi(\theta)$ is the tangent of the angle between a generator of the surface and the horizontal. In the deformed state, the sheet is in partial contact with the hoop so that it fits inside the axisymmetric envelope cone $z = \rho\epsilon$. If the sheet is very thin and the deformations are moderate ($\epsilon \leq 0.4$), the irreversible plastic deformation near the tip is small. Therefore we can use the theory of elasticity to describe its shape reasonably well. Away from the core the developable cone $\rho\psi(\theta)$ is determined by minimizing the bending energy subject to the constraints due to inextensibility and contact. Generalizing the functional given in ref. 11 to account for possibly large curvatures of the sheet requires the minimization of the following functional for $\psi(\theta)$:

$$L(\psi) = U_b + \lambda \int_{-\pi}^{\pi} \left[1 - \frac{(1 + \psi^2 + \psi'^2)^{1/2}}{1 + \psi'^2} \right] d\theta + \int_{-\pi}^{\pi} b(\theta)(\epsilon - \psi) d\theta \quad (1)$$

Here $U_b = (E_b/2) \ln(R_p/R_c) \int_{-\pi}^{\pi} [(\psi + \psi'')^2(1 + \psi^2)] / (1 + \psi'^2 + \psi''^2)^{5/2} d\theta$ is the bending energy, and E_b is the bending stiffness. The second term on the right enforces the inextensibility constraint and λ is proportional to the hoop stress. The third term enforces the requirement that the deformed sheet lies inside the perfect cone $z = \rho\epsilon$, with $b(\theta) = 0$ when $\psi \geq \epsilon$, and $b(\theta) \geq 0$ when $\psi = \epsilon$, and $b(\theta)$ related to the normal reaction of the supporting hoop on the conical dislocation. Figure 1 shows a numerically computed shape of the developable cone accounting for large bending deformations, and yields a solution valid everywhere except in the neighbourhood of the tip where the effects of stretching cannot be neglected.

We now estimate the core size characterized by the radius R_c which results from a balance between the bending and stretching energy. In the core of area ΔS , the stretching energy is $U_s \approx E_s \gamma^2 \Delta S$, where γ is the in-plane strain and E_s is the stretching stiffness. The bending energy in the core is $U_b \approx E_b \kappa^2 \Delta S$ where κ is the mean curvature in the core¹⁵. Equating these energies yields $U_b/U_s \approx (E_b/E_s)(\kappa/\gamma)^2 \approx O(1)$. The stretching strain can be estimated from the change in length of a typical generator of length R , the scale over which forces and torques are exerted. This generator is stretched to a length $\sqrt{R^2 + \epsilon^2 R_c^2}$; then the strain is $\gamma \approx (\sqrt{R^2 + \epsilon^2 R_c^2} - R)/R \approx (\epsilon R_c/R)^2$. The bending strain is simply the mean curvature κ which we now estimate. When $\psi, \psi' \ll 1$,

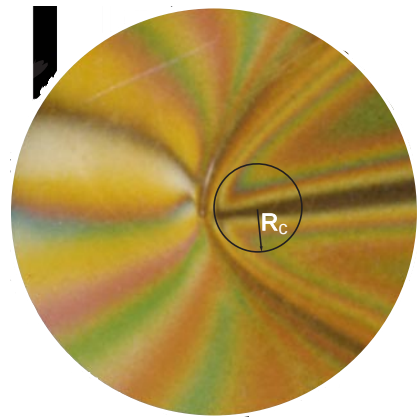


Figure 2 Geometry of a real conical dislocation. Cross-polarizers are used to view the reflected light from a painted sheet deformed into a conical dislocation. Isochromatic lines correspond to lines with no radial curvature, and coincide with generators of the cone away from the tip. In the core of the dislocation close to the tip, the sheet stretches into a surface of double curvature. Here the isochromatic lines have a radius of curvature R_c .

$U_b \approx \kappa^2 \approx \psi^2 \approx \epsilon^2$, while when $\psi, \psi' \gg 1$, $U_b \approx \kappa^2 \approx \psi^2 \psi^4 / \psi^5 \approx \epsilon$ (E.C. and L.M., manuscript in preparation). Therefore for small deformations, $\kappa \approx \epsilon/R_c$, while for large deformations, $\kappa \approx \epsilon^{1/2}/R_c$. This cross-over in the scaling occurs due to the change in the geometry of the sheet as ψ, ψ' are small or large compared to unity, so that the curvature (and the force) changes in response to it. Substituting in these strains into the energy balance yields the scaling law

$$R_c \approx \left(\frac{E_b}{E_s} \right)^{1/6} \epsilon^{-p} R^{2/3}; \quad p = \begin{cases} 1/3, & \epsilon \ll 1, \\ 1/2, & \epsilon \approx O(1) \end{cases} \quad (2)$$

consistent with the experimental results shown in Fig. 3a. This relation is valid for thin sheets of arbitrary materials subjected to large deformations, and extends the results obtained for the weak deformations of ridges⁸. The first factor in the scaling law is associated with material properties; for an isotropic material $E_b/E_s \approx h^2/(1 - \nu^2)$, where h is the sheet thickness and ν is Poisson's ratio. The second arises from the geometry of deformation as ϵ characterizes the cone angle. This factor also hides a subtle dependence on the force $F = \partial U_b / \partial \epsilon$ since the force-deflection relation, discussed later, is of the form $\epsilon = \nu(FR/E_b)$, where $\nu(s)$ is a dimensionless function. There is also a logarithmic dependence on the sheet radius R_p but it is difficult to detect experimentally. The third factor characterizes the length R associated with the moment arm of the reaction force along the hoop.

Next we consider the behaviour of the conical dislocation when a force F is applied along the axis of the supporting hoop. In our experiments, a round tip of diameter 0.5 mm is attached to a load control cell which is used to measure the applied force. Care is taken to minimize frictional effects by lubricating the contact zone between the sheet and supporting hoop. We restrict our range of deformations to minimize inelastic effects. In Fig. 3b we plot F versus ϵ and see that for $\epsilon \leq 0.1$, $F \propto \epsilon$ but for $\epsilon \geq 0.1$ $F \approx \epsilon^p$, $p < 1$. The location of this cross-over coincides with that in Fig. 3a for the cross-over in the scaling law for R_c , suggesting that it is geometric effect. As ϵ increases, the effective reaction force from the hoop, which is normal to the sheet in the absence of friction, starts to become more horizontal, while the torque increases. This decreases the resistance to further deformation and softens the system, eventually leading to an instability. We quantify this by evaluating F in terms of the bending energy U_b , $F = \partial U_b / \partial d = (E_b/R) \ln(R_p/R_c) \partial u(\epsilon) / \partial \epsilon$. Here $u(\epsilon)$ is a dimensionless function such that $u \approx \epsilon^2$ for small ϵ and $u \approx \epsilon$ for large ϵ , following an argument identical to that for the scaling of the curvature. We use

experimentally determined values of R_p and E_b for each material, and the linear part of the experimental curve to estimate R_c . A numerical method is used to solve the boundary-value problem associated with the Euler–Lagrange equations that arise from minimizing the functional $L(\psi)$ in equation (1). We use the analytical solution determined in ref. 11 as an initial guess for small ϵ and then use an iterative procedure to determine the complete $F - \epsilon$ curve shown in Fig. 3b. The experimental data are well explained by our theory that accounts only for bending, and indicate that stretching is not important in this case. Continuing this computation beyond the experimentally determined points, we find that the stiffness of the system eventually goes to zero, signifying a subcritical (snap-through) dynamic instability in a force-controlled experiment¹⁶ that is arrested when the sheet contacts itself.

So far we have described the structure and the response of a single centrally forced sheet that leads to a conical dislocation with a fold that is rotation-invariant. In the course of crumpling, two such dislocations often interact with each other, forming ridges that break this rotation symmetry. Prior studies⁸ have focused on the

structure and linear response of these ridges, but have artificially excluding the effects of the conical dislocations that they connect. We now consider them explicitly in the context of an experiment by forcing an acetate sheet into a circular hoop at two non-central locations. The two probes lie on a diameter and are separated by a distance $2D$, so that we have an additional control parameter $\delta = D/R$. As the probes are moved axially by a distance d , two conical dislocations appear at symmetrical locations with respect to the centre of the hoop. In Fig. 4a we show one such configuration, computed by solving the Euler–Lagrange equations obtained by minimizing $L(\psi)$ in equation (1). This solution minimizes the bending energy of the sheet but does not account for stretching, so that it leads to a flat interlying region instead of a stretched ridge in between the dislocations. To compare these calculated results with experiment, we consider the mechanical response of this structure characterized by a $F - \epsilon$ curve, shown in Fig. 4b. The theoretical curve has no tunable parameters, and shows that bending alone accounts very well for the measured force. To quantify this further, we recall that the stored bending energy in this structure is $U_b = E_b \epsilon^2 f(\delta, \epsilon) \ln(R_p/R_c)$, with $f(\delta, \epsilon) \approx 84.7\epsilon^2(1 + \text{const.}\delta)$ for

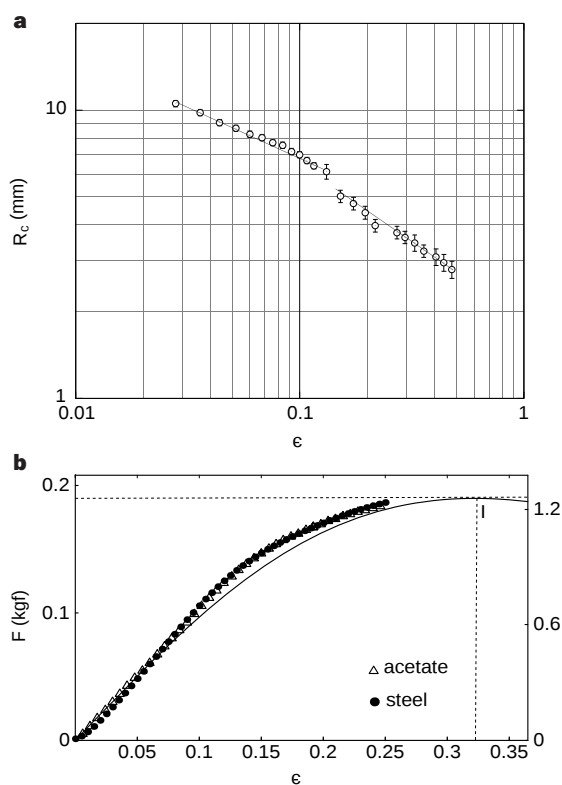


Figure 3 Mechanical response of a conical dislocation. **a**, The core size R_c as a function of the deformation of the conical dislocation characterized by $\epsilon = d/R$. In our experiments we used a series of thin sheets of steel (bending stiffness $E_b = 7.76 \times 10^{-3}$ N m, thickness $h = 0.075$ mm) and acetate (bending stiffness $E_b = 5.1 \times 10^{-4}$ N m, $h = 0.1$ mm). The radii of the sheets varied from 15 to 90 mm, while the radius of the supporting hoop was 5 mm less than the radius of the sheet. The transverse displacement d of the centre of the thin sheet relative to the plane of the hoop is measured using a precision micrometer with an accuracy of $10 \mu\text{m}$. The data are best fitted by the power laws $R_c \propto \epsilon^{-0.33 \pm 0.01}$, $\epsilon \leq 0.1$, $R_c \propto \epsilon^{-0.50 \pm 0.02}$, $\epsilon \geq 0.1$ (solid lines). The jump in the data at $\epsilon \approx 0.1$ corresponds to using two different hoop sizes. The error bars are larger as ϵ increases due to a decrease in the core size which leads to having fewer data points. **b**, The force–deformation curve for a conical dislocation made of acetate (open triangles; left ordinate) and steel (filled circles; right ordinate), compared with the theoretical curve (solid line) calculated using the expression for the force F given in the text, with no adjustable parameters. We ascribe the small systematic overshoot of the experimental data to frictional effects. The maximum F in the theoretical curve corresponds to the location of a subcritical (snap-through) instability in a force-controlled experiment.

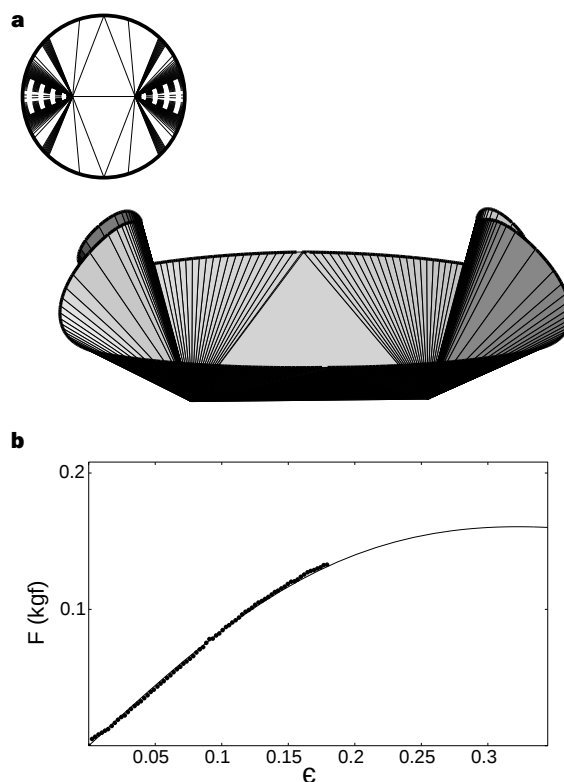


Figure 4 Geometry and mechanical response of two interacting conical dislocations. **a**, When a sheet is forced transversely into a rigid circular hoop at two locations along a diameter, it deforms into a boat-like structure with two symmetrically placed conical dislocations. This solution with least energy is calculated by solving the Euler–Lagrange equations that arise from minimizing $L(\psi)$ in equation (1), starting with an analytical solution for two weakly deformed conical dislocations with $\delta = D/R = 0$, $\epsilon = d/R \ll 1$, and using an iterative continuation scheme until $\delta = 0.38$, $\epsilon = 0.345$. This solution accounts only for the effects of bending, and the associated generators are shown at top left. The diamond-shaped region between the dislocations consists of two flat planes, and is separated from the conical regions by lines along which the curvature of the sheet suffers a discontinuity. Accounting for stretching and bending will lead to boundary layers at these locations. **b**, The force–deformation curve for two interacting conical dislocations in an acetate sheet, compared with the theoretical curve (solid line), with no adjustable parameters. We see that a theory accounting for bending alone is able to explain the response of this structure which exhibits a sub-critical instability for moderate ϵ , as evidenced by the presence of a maximum in the $F - \epsilon$ curve.

small ϵ , δ while the energy which resides in the stretched ridge⁸ is $U_s = E_b \epsilon^{7/3} (2D/h)^{1/3}$, so that $U_s/U_b \approx 0.05$ for typical experimental parameters. These results indicate that in the early stages of crumpling, when large deformations of a few conical dislocations are much more likely to occur, bending dominates stretching. Finally, by continuing the computation of the $F - \epsilon$ curve beyond the experimentally accessible parameters, we see the appearance of a snap-through instability similar to that for a single dislocation.

A series of such events (that is, geometric softening—dynamic snap-through—local topological stiffening) provides us with a microscopic mechanism for the crumpling of a large thin elastic sheet. As the sheet is deformed by a force, it forms a developable cone that deforms, softens and eventually becomes dynamically unstable, and an acoustic pulse is emitted when the sheet pops into a folded configuration. This stiffens the sheet locally, but soon new developable cones (and stretched ridges which connect them) begin to form. Ridges may buckle in two ways. (1) In the plane of the ridge by forming a developable cone about which the ridge pivots and folds, locally leading to roughly the same scenario as that for a single conical dislocation, or (2) in a direction perpendicular to the ridge by forming two dislocations that move apart along a new ridge about which the original ridge folds, as seen during the bending of a drinking straw. A cascade of these instabilities on ever-decreasing length scales leads to the formation of new conical dislocations as the sheet crumples, and the energy of deformation is pumped down to smaller and smaller scales. As the size of these folds becomes smaller, the incremental deformation is concomitantly less, and a cross-over to the regime where stretching and bending deformations are of the same order is likely⁸. However any analysis of this stage in crumpling must also account for inelastic deformations. On length scales much larger than the thickness but much smaller than the length or breadth of the sheet, these dynamical snap-throughs constitute a self-similar cascade and are accompanied by acoustic emissions. While preliminary experiments^{17,18} are suggestive of power-law behaviour for the statistics of these sounds, they remain incompletely quantified: much work remains to be done in this area. □

Received 1 February; accepted 17 June 1999.

- Calladine, C. *Theory of Shell Structures* (Cambridge Univ. Press, Cambridge, 1983).
- Wierzbicki, T. & Jones, N. (eds) *Structural Failure* (Wiley Interscience, New York, 1988).
- Connelly, R. Rigidity and energy. *Invent. Math.* **66**, 11–33 (1982).
- Nelson, D. R., Piran, T. & Weinberg, S. *Statistical Mechanics of Membranes and Surfaces* (World Scientific, Singapore, 1988).
- Amirbayat, J. & Hearle, J. W. S. The complex buckling of flexible sheet materials. *Int. J. Mech. Sci.* **28**, 339–370 (1986).
- da Vinci, Leonardo *Notebooks Vol. I, Studies of Drapery* (Dover Reprint, New York, 1984).
- Rayleigh, Lord, *Theory of Sound Vol. I, Ch. X a* (Dover, New York, 1945).
- Lobkovsky, A., Gentges, S., Li, H., Morse, D. & Witten, T. Stretched ridges in crumpling. *Science* **270**, 1482–1485 (1995).
- Lobkovsky, A. & Witten, T. A. Properties of ridges in elastic membranes. *Phys. Rev. E* **55**, 1577–1589 (1997).
- Chaieb, S. & Melo, F. Experimental study of developable cones. *Phys. Rev. Lett.* **80**, 2354–2357 (1998).
- Cerda, E. & Mahadevan, L. Conical surfaces and crescent singularities in crumpled sheets. *Phys. Rev. Lett.* **80**, 2358–2361 (1998).
- Ben Amar, M. & Pomeau, Y. Crumpled paper. *Proc. R. Soc. Lond. A* **453**, 729–755 (1997).
- Nabarro, F. R. N. *Theory of Crystal Dislocations* (Dover, New York, 1993).
- Struik, D. J. *Lectures on Classical Differential Geometry* (Dover, New York, 1988).
- Love, A. E. H. *A Treatise on the Mathematical Theory of Elasticity* (Dover, New York, 1944).
- Timoshenko, S. & Gere, J. *Theory of Elastic Stability* (McGraw-Hill, New York, 1961).
- Kramer, E. & Lobkovsky, A. Universal power law in the noise from a crumpled elastic sheet. *Phys. Rev. E* **53**, 1465–1468 (1996).
- Houle, P. & Sethna, J. Acoustic emission from crumpling paper. *Phys. Rev. E* **54**, 278–283 (1996).

Acknowledgements

E.C. was supported by the Chilean Presidente de la República postdoctoral fellowship during the course of this work at MIT in 1997–98. S.C. was supported by a postdoctoral fellowship at Universidad de Santiago de Chile in 1997–98 during the course of this work. Additional support was provided by the Chilean Cátedra Presidencial en Ciencias (F.M.), the Karl van Tassel career development chair and the Sloan fund (L.M.) at MIT.

Correspondence and requests for materials should be addressed to L.M. (l_m@mit.edu).

Direct observation of *d*-orbital holes and Cu–Cu bonding in Cu₂O

J. M. Zuo*, M. Kim*, M. O'Keeffe† & J. C. H. Spence*

* Department of Physics and Astronomy, † Department of Chemistry, Arizona State University, Tempe, Arizona 85287, USA

A striking feature of metal oxide chemistry is the unusual electronic and chemical behaviour of Cu(I) and Ag(I): a case in point is that detailed understanding of Cu–O bonding is essential to the theory of high-temperature copper oxide superconductors. Both cations are usually coordinated in a linear fashion to two oxygens, particularly for Cu(I). In many compounds, the Cu(I) and Ag(I) cations also adopt close-packed (and related) configurations with short metal–metal distances that are strongly suggestive of the occurrence of metal–metal bonding^{1,2} despite their formal nd^{10} configuration. Such observations have been explained^{3,4} by invoking the participation in bonding of electronic orbitals of higher principal quantum number—that is, $(n+1)s$ and $(n+1)p$ —accompanied by the creation of *d*-orbital holes on the metal ion. To test this hypothesis, we have used a recently developed method of quantitative convergent-beam electron diffraction⁵ combined with X-ray diffraction to map the charge-density distribution in the simple oxide Cu₂O, the results of which we then compare with electronic-structure calculations. We are able to image directly the *d* holes on the copper atoms, and also demonstrate the existence of Cu–Cu bonding in this compound.

Cu₂O has a cubic structure with no free internal parameters (only Ag₂O is isostructural). The copper atoms are at the points of a face-centred-cubic lattice, with oxygen atoms in tetrahedral sites at positions (1/4,1/4,1/4) and (3/4,3/4,3/4) of the cubic cell. The resulting arrangement of Cu–O links is made up of two interpenetrating networks (Fig. 1). The simplest description of Cu₂O using an ionic model with closed-shell (spherical) Cu⁺ and O²⁻ ions is known to be inadequate. Not only does it fail to explain the observed linear 2-coordination of Cu but also it is not in accord with the observation that the two sublattices repel each other electrostatically, so that to account for their interpenetration some short-range Cu–Cu attractive interaction must be invoked⁶. We note that the closest approach of atoms of the two networks is a Cu–Cu distance of 3.02 Å—the shortest O–O distances are 3.70 Å.

Considerable progress has been made in mapping the charge density of light-element molecular crystals by X-ray diffraction⁷, especially using a synchrotron source⁸. The extension of this method

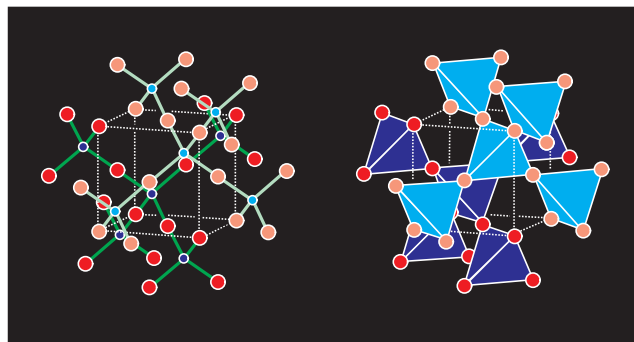


Figure 1 The structure of Cu₂O. Left, as a ball and stick model with O atoms blue, Cu atoms red and bonds green. One network is coloured darker than the other. Note that there are no bonds joining the two nets. Right, as corner-connected OCu₄ tetrahedra. Dark and light tetrahedra are on independent networks. In both sketches, dotted white lines outline a unit cell.

to inorganic solids of interest to material science is complicated by the X-ray extinction effect, which masks the much smaller bonding effect at low scattering angles, especially for heavy elements⁹. The electron-diffraction method used here allows the extinction-free measurement of low-order crystal structure factors, which are most sensitive to valence electrons. Following an earlier suggestion¹⁰, we combine these structure factors with X-ray diffraction results for weak and high-order reflections. Electron diffraction was also used

to assess the error due to extinction in X-ray diffraction data, and to replace strongly affected reflections with electron-diffraction data. The result clearly shows that ambiguities in previous X-ray studies^{11,12} of the charge density of Cu₂O (cuprite) are due to large extinction effects on the strong low-order reflections.

Accurate measurements of the low-order structure factors were made with a versatile, quantitative, convergent-beam electron diffraction (CBED) technique³ that we have developed recently. The method takes advantage of the small probe (of nanometre dimensions) available in our electron microscope. Using this probe, one can almost always find a region of perfect crystals, for which the perfect-crystal theory of dynamical diffraction can be applied. The measurements are made by comparing experimental intensity profiles across CBED disks (rocking curves) with calculations, using a goodness of fit (GOF) criterion, as illustrated in Fig. 2. The intensity is calculated using the Bloch wave method, with structure factors, absorption coefficients, the beam direction and thickness treated as refinable parameters. Structure factors for the (531) and higher-order reflections out to (14,4,2) were taken from X-ray measurements¹¹. Weak (odd, even) and very weak (even, even, odd) reflections—except (110)—were also taken from X-ray work.

Multipole refinement^{7,13} provides the most efficient parametrization of real space charge density, and gives a result which is insensitive to the missing reflections in the collected data set¹¹. In this method, the crystal charge density is fitted by a sum of non-spherical pseudo-atomic densities. These consist of a spherical-atom (or ion) charge density obtained from multi-configuration Dirac-Fock calculations¹⁴ with variable orbital occupation factors to allow for charge transfer, and a small non-spherical part in which local symmetry-adapted spherical harmonic functions were used. In addition, atomic vibrations were accounted for using the Charlier–Gram expansion⁷ for the temperature factor; as the electron and X-ray measurements were done at different temperatures, the combined refinement used the almost linear dependence of temperature factors between 100 and 300 K (ref. 12). Refinements with, and without, anharmonic terms in the temperature factor clearly show the importance of an anharmonic term for Cu, especially for high-order reflections with $s = \sin \theta / \lambda > 1.0 \text{ \AA}^{-1}$. In either case the charge transfer from Cu to O refined to 1.01(5); that is, Cu⁺ and O²⁻.

Compared to previous results^{11,12}, we obtain much better estimates for the non-spherical distortion of O and Cu. This improvement is due to the absence of extinction effects in our electron-diffraction measurements. X-ray diffraction has a particularly strong extinction effect for strong reflections with small s values. For example, for Cu₂O(200), the X-ray structure factor $F^X \approx 78.8$ electrons per cell, and the measured value by X-ray diffraction is

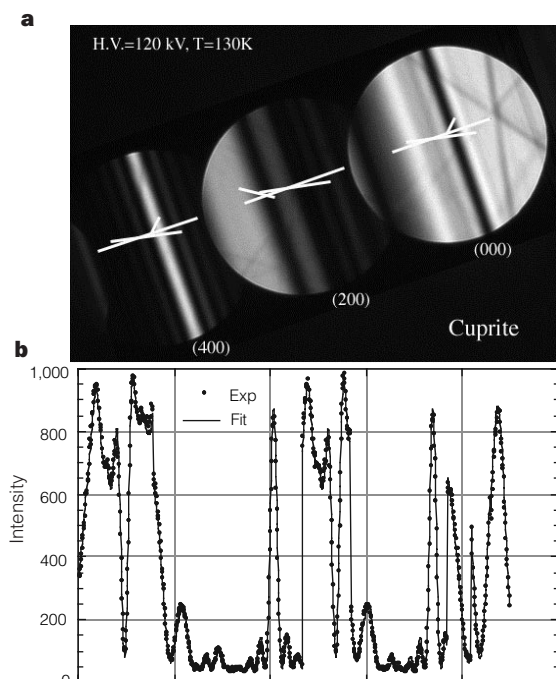


Figure 2 An example of electron diffraction structure factor measurement for Cu₂O (200) and (400). **a**, Experimentally recorded and processed CBED pattern; **b**, best fit for points along lines indicated in **a**, displayed sequentially along the x -axis and with the recorded intensity on the y -axis in counts. The experiment was performed using an LEO-912 Ω energy-filtering electron microscope with a Gatan liquid-nitrogen-cooled sample holder. The specimen used is a natural Cu₂O (cuprite) cooled to ~ 130 K. A 15-eV energy-filtering slit was placed around the zero-loss peak to remove the contribution from inelastically scattered electrons, which form a background due to plasmon and other loss processes. Off-zone-axis systematic diffraction conditions were used to collect diffraction intensities for reflections up to (400). The CBED patterns were recorded on both a slow-scan CCD camera and on Fuji imaging plates, which were then processed for fitting⁵. The refinement was carried out using the EXTEL program⁵.

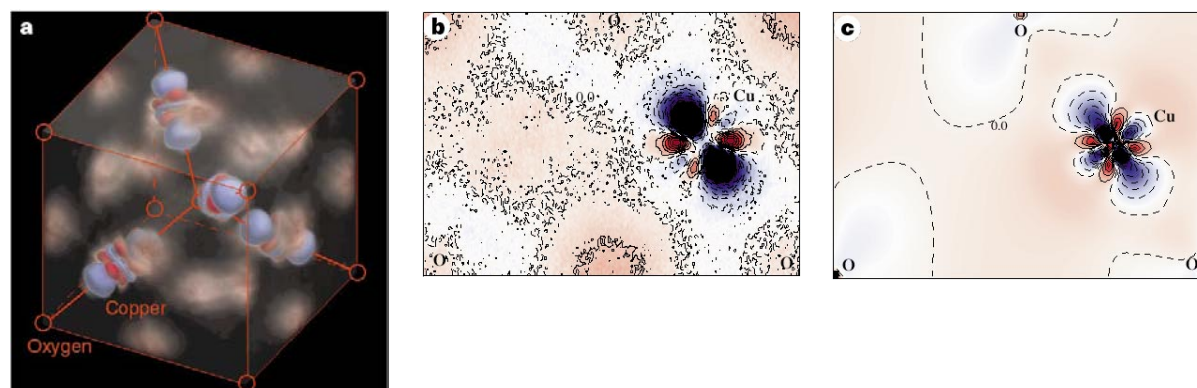


Figure 3 The experimental and theoretical difference maps between the static crystal charge density and superimposed spherical O²⁻ and Cu⁺ ions. **a**, A three-dimensional rendering of the experimental difference map from the multipole model fitting with anharmonic temperature factors. The map was made using the T3D program (Fortner Research, USA) using a colour scheme of blue ($\Delta\rho < 0$), white ($\Delta\rho = 0$) and red

($\Delta\rho > 0$). A translucency factor was used to remove the mostly white background.

b, Contour map of the difference charge density in a (110) plane with a contour increment of 0.2 electrons per $\text{\AA}^2$ and coloured using the same colour scheme as in **a**. The dashed line is for contours with $\Delta\rho \leq 0$. **c**, The difference charge density map obtained from the FLAPW (LDA) calculations, plotted in the same way as **b**.

about 90% of the actual value, with a difference of ~ 7.9 electrons per cell. For this reflection, the difference between the ionic and the neutral-atom models is about 0.07 electrons per cell. To detect this very small difference, the extinction correction used has to have better than 1% accuracy. By comparison, the electron-diffraction measurement has an accuracy of 0.04 electrons per cell. The electron measurements allow us to evaluate the extinction effect in the X-ray data and exclude those with a large extinction factor.

The large non-spherical contributions in the multipole fitting point to an appropriate description of the charge density in terms of distorted ions; the effect of covalent bonding interactions (that is, interactions other than the electrostatic between spherical ions) can be seen by comparing the 'measured' (multipole refinement) electron density with that resulting from a model consisting of superposition of spherical ions. In Fig. 3a we show a three-dimensional plot of the difference between the static crystal charge density (obtained from the multipole fitting to experiment) and the superimposed charge density of spherical O^{2-} and Cu^+ ions calculated by the multi-configuration Dirac-Fock method. The charge density of the O^{2-} ion was calculated using a Watson sphere of 1.2 Å radius; this is the value that gives the best fit to experiment within the limitation of a spherical ion model. Figure 3b shows a contour map of the difference charge density for the (110) plane; this was synthesized by introducing random noise to the Fourier coefficients for each pixel, to illustrate the uncertainty in the mapped charge density. The noise introduced reflects the uncertainty in the estimated multipole parameters. For comparison, in Fig. 3c, we also plot the theoretical difference charge density map obtained from the full potential linearized augmented plane wave (FLAPW) method, using the local density approximation (LDA) for exchange and correlation and the WIEN95 program¹⁵ (this is an updated version of the program in ref. 16). This calculation improves on the previous calculation¹⁷ using the muffin-tin potential approximation. For the theoretical map, the charge densities of O^{2-} and Cu^+ ($3d^{10}$) ions were calculated using the same LDA as the FLAPW calculation.

The electron density difference shown in Fig. 3a would be zero everywhere if cuprite were purely ionic (that is, if it consisted of spherical ions). The difference, here seen unambiguously for the first time (to our knowledge), confirms earlier theoretical speculation¹⁻⁴ that a covalent contribution exists. The correspondence between our experimental map and the classical diagrams of d_z^2 orbitals sketched in textbooks is striking. All our difference maps show strong non-spherical charge distortion around the copper atoms, with the characteristic shape of d orbitals, and excess charge in the interstitial region. There is little variation around oxygen in both the experimental and the theoretical results, which suggests that an O^{2-} anion description is valid. The most significant difference between experiment and theory is around Cu, and the charge in the interstitial region. The charge modification around Cu in the experiment is broader and larger than the theory. The experimental map also shows a large (~ 0.2 electrons per Å³) positive peak in the unoccupied tetrahedral interstitial region of the four neighbouring Cu atoms, which suggests a strong Cu^+-Cu^+ covalent bonding. For comparison, the Si-Si covalent bond peak is about 0.21 electrons per Å³ (ref. 18). We note that these differences

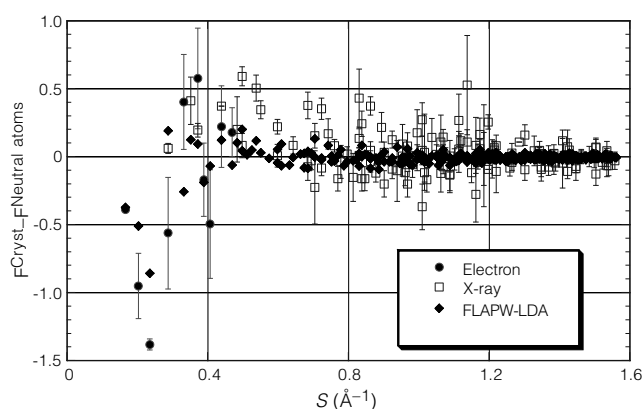


Figure 4 Comparison between experimental and theoretical structure factors (F). The structure factor shown is the difference between the crystal and the neutral atoms model. The error bar of first reflection is smaller than the circle.

between experiment and theory are experimentally significant. In Fig. 4, we compare experimental and theoretical structure factors. The experiment structure factor clearly shows a large deviation from that of superimposed neutral atoms. The difference between theory and experiment is generally larger than the experimental error.

The non-spherical charge density around Cu^+ can be interpreted as being due to the hybridization of d electrons with higher-energy unoccupied s and p states^{3,4,17}. Among these states, hybridization is only allowed for d_z^2 and $4s$ by symmetry, and when this happens part of the d_z^2 state becomes unoccupied (a d -orbital hole). These states are responsible for the spatial distribution of the deficiency in the map shown in Fig. 3a. The complementary empty states are important for spectroscopies which probe empty states, such as electron energy-loss spectroscopy and X-ray absorption spectroscopy¹⁹. The experimental studies reveal that the unoccupied states are predominately $Cu-d$ character for the Cu $L_{2,3}$ edge; theory shows that they originate from hybridized d_z^2 orbitals. This theoretical interpretation, based on the calculated partial density of states of the one-electron band structure, is supported by the generally good agreement with experimental spectroscopy of both occupied and unoccupied states¹⁹. From the characteristic and localized distortion around Cu, we can estimate the population of the outer valence shells from the multipole model coefficients^{7,11}. From the data in Table 1, we estimate the hybridization coefficient, $|x|$, between d_z^2 and $4s$ to be 0.36, so that about 0.22 electrons per atom are removed from d_z^2 states. This is in good agreement with the calculated density of states, which gives a similar number for x . □

Received 12 March; accepted 16 July 1999.

- Jansen, M. Homoatomic $d^{10}-d^{10}$ interactions: their effects on structure and chemical and physical properties. *Angew. Chem. Int. Edn. Engl.* **26**, 1098–1110 (1987).
- Pyykkö, P. Strong closed-shell interactions in inorganic chemistry. *Chem. Rev.* **97**, 597–636 (1997).
- Orgel, L. E. Stereochemistry of metals of the B sub-group. Part I, ions with filled d -electron shells. *J. Chem. Soc.* 4186–4190 (1958).
- Merz, K. M. & Hoffman, R. $d^{10}-d^{10}$ interactions: multinuclear copper(I) complexes. *Inorg. Chem.* **27**, 2120–2127 (1988).

Table 1 Cu_2O multipole model refinement results

O			Cu			q	R	
α (Å ⁻¹)	P_{oct}	P_{hex}	α (Å ⁻¹)	P_{20}	P_{40}	(e)	(%)	
2.24 (10)	-1.65 (14)	3.28 (10)	15.7(10)	-0.195 (30)	-0.219 (38)	-0.043 (20)	1.01 (5)	0.71

In the multipole model, crystal charge density is a sum of aspherical 'atoms': $\rho^s(\mathbf{r}) = \rho^s(\mathbf{r}) + \sum_{lm} P_{lm} N_{lm} R_l(r) Y_{lm}(\theta, \phi)$. Here ρ^s is the spherical charge with $\rho_{Cu}^s = \rho_{Cu}^s + (1 - q)(\rho_{Cu}^s - \rho_{Cu}^s)$, and $\rho_O^s = \rho_O^s + q(\rho_{O^{2-}}^s - \rho_O^s)$ with q as the charge transfer from Cu to O; $R_l(r) = r^l \exp(-\alpha r)$ ($l_3 = 3$; $l_4 = 4$) is the radial function with population coefficient P_{lm} ; N_{lm} is the density-normalization coefficient. Numbers in parentheses are estimated standard deviations.

- Zuo, J. M. Quantitative convergent beam electron diffraction. *Mater. Trans. JIM* **39**, 938–946 (1998).
- O'Keeffe, M. Madelung constants for the C3 and C9 structures. *J. Chem. Phys.* **38**, 3035–3038 (1963).
- Coppens, P. *X-ray Charge Densities and Chemical Bonding* (Oxford, New York, 1997).
- Koritsanszky, T. *et al.* Accurate experimental electronic properties of DL-proline monohydrate obtained within 1 day. *Science* **279**, 356–358 (1998).
- Stewart, R. F. & Spackman, M. in *Structure and Bonding in Crystals* (eds O'Keeffe, M. & Navrotsky, A.) 279–298 (Academic, New York, 1981).
- Smart, D. J. & Humphreys, C. J. The application of electron diffraction to determining bonding charge densities in crystals. *Int. Phys. Conf. Ser.* **52**, 211–214 (1980).
- Restori, R. & Schwarzenbach, D. Charge density in cuprite, Cu_2O . *Acta Crystallogr. B* **42**, 201–208 (1986).
- Kirfel, A. & Eichhorn, K. Accurate structure analysis with synchrotron radiation, the electron density in Al_2O_3 and Cu_2O . *Acta Crystallogr. A* **46**, 271–284 (1990).
- Stewart, R. F. Electron population analysis with rigid pseudoatoms. *Acta Crystallogr. A* **32**, 565–574 (1976).
- Rez, D., Rez, P. & Grant, I. Dirac-Fock calculations of X-ray-scattering factors and contributions to the mean inner potential for electron-scattering. *Acta Crystallogr. A* **50**, 481–497 (1994); correction **53**, 522 (1997).
- Blaha, P., Schwarz, K. & Luitz, J. WIEN97 (Vienna Univ. of Technol., Vienna, 1997).
- Blaha, P., Schwarz, K., Sorantin, P. & Trickey, S. B. Full-potential, linearized augmented plane-wave programs for crystalline systems. *Comput. Phys. Commun.* **59**, 399–415 (1990).
- Marksteiner, P., Blaha, P. & Schwarz, K. Electronic structure and binding mechanism of Cu_2O . *Z. Phys. B* **64**, 119–127 (1986).
- Zuo, J. M., Blaha, P. & Schwarz, K. The theoretical charge density of silicon: experimental test of exchange and correlation potentials. *J. Phys. Condens. Matt.* **9**, 7541–7561 (1997).
- Ghijzen, J. *et al.* Electronic structure of Cu_2O and CuO . *Phys. Rev. B* **38**, 11322–11330 (1988).

Acknowledgements

We thank A. Avilov for discussions, and D. Schwarzenbach for providing the X-ray data. This work was supported by the NSF.

Correspondence and requests for materials should be addressed to J.M.Z. (e-mail: Zuo@asu.edu).

Discerning vibronic molecular dynamics using time-resolved photoelectron spectroscopy

Valérie Blanchet, Marek Z. Zgierski, Tamar Seideman & Albert Stolow

Steacie Institute for Molecular Sciences, National Research Council of Canada, Ottawa, Ontario, Canada K1A 0R6

Dynamic processes at the molecular level occur on ultrafast time scales and are often associated with structural as well as electronic changes. These can in principle be studied by time-resolved scattering^{1–3} and spectroscopic methods, respectively. In polyatomic molecules, however, excitation results in the rapid mixing of vibrational and electronic motions, which induces both charge redistribution and energy flow in the molecule^{4,5}. This 'vibronic' or 'non-adiabatic' coupling is a key step in photochemical⁶ and photobiological processes⁷ and underlies many of the concepts of molecular electronics⁸, but it obscures the notion of distinct and readily observable vibrational and electronic states. Here we report time-resolved photoelectron spectroscopy measurements that distinguish vibrational dynamics from the coupled electronic population dynamics, associated with the photo-induced internal conversion, in a linear unsaturated hydrocarbon chain. The vibrational resolution of our photoelectron spectra allows for a direct observation of the underlying nuclear dynamics, demonstrating that it is possible to obtain detailed insights into ultrafast non-adiabatic processes.

The Born-Oppenheimer approximation (BOA), an adiabatic separation of electronic from nuclear motions, has the pivotal role in defining the potential energy surface and thus permits a mechanistic picture of molecular dynamics. The breakdown of the BOA is due to the motions of the atoms near the intersections of

potential surfaces belonging to different electronic states. This mixing of vibrational and electronic motions in excited molecules can in principle be inferred from both frequency-resolved and time-resolved optical spectroscopies. Time-resolved experiments may be generally understood as preparing an excited state with a pump pulse and then, as a function of time, projecting it with a probe pulse onto a final state which acts as a template^{9–17}. Ideally, the final state should be well characterized. Photoionization detection has several conceptual and practical advantages¹⁰. The ion state may be well characterized by independent methods such as high resolution photoelectron spectroscopy and *ab initio* computation. Photoelectron detection has the additional advantage that the signal may be dispersed with respect to kinetic energy and angular distribution. Photoelectron spectroscopy has, compared with optical spectroscopies, fairly relaxed selection rules: any molecular state can be ionized (no 'dark' states). A simplified but very useful picture is that emission of an independent outer electron can occur without the simultaneous electronic reorganization of the ion core (Koopmans' approximation). Partial photoionization probabilities can differ drastically with respect to the molecular orbital nature of the neutral electronic state. If a given electronic configuration correlates—upon removal of the outermost electron—to the electronic configuration of the ground electronic state of the cation, then the corresponding photoionization probability is much higher than if it does not.

Time-resolved photoelectron spectroscopy (TRPES) has been proposed as a technique for the study of ultrafast non-adiabatic processes^{10–14} and spin-orbit coupling¹⁵ in isolated polyatomic molecules. The TRPES method has been shown to provide details

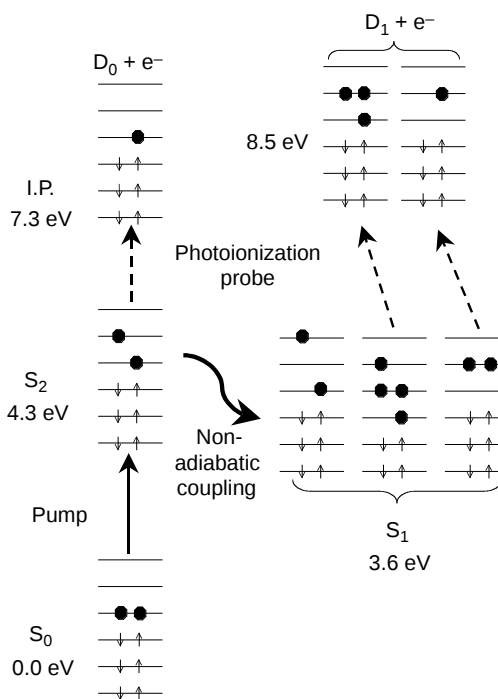


Figure 1 Molecular orbital configurations and photoionization correlations in all-trans 2,4,6,8-decatetraene, DT. The optically coupled S_2 (excitation at 4.3 eV shown by straight solid arrow) state is a singly excited configuration, whereas the lower-lying optically forbidden S_1 (3.6 eV) state is composed of both singly and doubly excited configurations. The curved solid arrow indicates non-adiabatic coupling in the excited state. On removal of the highest-lying electron, S_2 correlates with the D_0 cation ground-state configuration with an ionization potential (I.P.) of 7.3 eV, whereas S_1 correlates predominantly with the D_1 cation excited-state configurations at 8.5 eV. It is therefore expected, assuming a single active electron, that the photoionization electronic channel should switch from $D_0 + e^-$ to $D_1 + e^-$ during the $S_2 \rightarrow S_1$ internal conversion. The dashed arrows indicate the most probable ionization channels.

of vibrational dynamics in simple molecules on single potential surfaces^{10,16,17}. In polyatomic molecules, non-adiabatic coupling leads to an evolving electronic state symmetry and therefore a changing correlation upon ionization. A switching of the photoelectron band is expected, providing a record of the electronic population dynamics. The detailed vibrational structure of each band simultaneously reveals the nuclear dynamics in each of the coupled electronic states.

We demonstrate our method by studying ultrafast internal conversion in a linear polyene: all-*trans* 2,4,6,8 decatetraene (DT, C₁₀H₁₄). Linear polyenes are hydrocarbon chains that have long been an area of fundamental and applied research¹⁸. Their non-adiabatic dynamics leads to the fundamental process of *cis-trans* photoisomerization. Polyenes form the light-harvesting antennae in vision (rhodopsin) and light-driven transmembrane proton pumps (bacteriorhodopsin). The investigation of polyene photophysics is central to our understanding of electron delocalization and electron correlation in molecules. They have also been used to test quantum chemical theory, as their lowest excited state contains doubly excited configurations¹⁹. DT provides a classic example of internal conversion in a polyene²⁰, making it a good candidate for testing our method.

The molecular orbitals that are important in the internal conversion of DT (ionization potential 7.3 eV) can be seen in Fig. 1. These calculated excited electronic states of the molecule and radical cation were obtained with the QCFF/PI + CISD (quantum consistent force field/ π -electrons + configuration interaction singles and doubles) method^{21,22}. The S₀(1¹A_g) electronic ground state is a single configuration. The first optically allowed transition, at 4.3 eV, is S₂(1¹B_u) $\leftarrow$ S₀. The S₂ state is a singly excited configuration. The lowest excited state, however, is the dipole forbidden S₁(2¹A_g) state²³ which arises from configuration interaction between singly and doubly excited A_g configurations, as shown in Fig. 1. Non-adiabatic coupling, leading to ultrafast internal conversion from S₂ to S₁, is promoted by vibrational motions of B_u symmetry. The ionization energies of the continua discussed here are low enough to ensure that the electrons are independent²⁴ and that Koopmans' approximation applies. The most probable electronic configurations of the cation expected on removal of the highest-lying electron from the neutral excited state are indicated in Fig. 1 by the dashed arrows. It can be seen that the S₂ excited state correlates with the D₀(1²B_g) ground electronic state of the cation. The S₁ state, by contrast, correlates predominantly with the D₁(1²A_u) first excited ionic state. We can expect, therefore, a change in the ion state correlation to occur during internal conversion.

Our study of the S₂ to S₁ internal conversion in DT used the femtosecond pump-probe TRPES method described previously^{10,11}. In Fig. 2a we show the energy level scheme relevant to this experiment. A femtosecond pump pulse at 287 nm prepared the excited S₂ state at its electronic origin. The system then evolves into a vibrationally hot (0.7 eV) S₁ electronic state via internal conversion. The evolving electronic symmetry is observed by projecting the wavepacket onto several cation electronic states using an ultraviolet probe photon of sufficient energy (here, 235 nm). As the non-adiabatic dynamics proceeds, the changing electronic character alters the photoionization electronic channel, leading to a band switching in the time-resolved photoelectron spectrum.

The experimental photoelectron kinetic energy spectra (Fig. 2b) are characterized by a rapid shift of electrons from an energetic band ($\epsilon_1 = 2.5$ eV) to a broad, structured low-energy band (ϵ_2). This shift is the direct signature of the changing electronic symmetry induced by non-adiabatic coupling. The ϵ_1 band is due to ionization of the S₂ into the D₀ ion state. The ϵ_2 band arises from photoionization of the S₁ state that correlates with the D₁ ion state: its appearance is due to population of the S₁ state by internal conversion. Integration of the photoelectron bands directly reveals the S₂ to S₁ internal conversion timescale of 386 ± 65 fs. A previous estimate²⁵ based on linewidth measurements was around 0.25 ps.

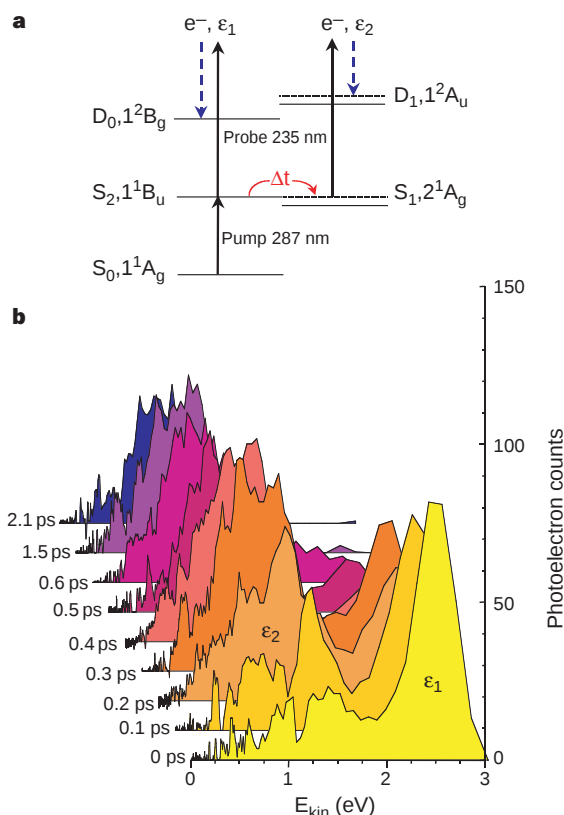


Figure 2 Time-resolved vibrational and electronic dynamics during internal conversion in DT. **a**, Level scheme in DT for one-photon probe ionization. The pump laser prepares the optically bright state S₂. Due to ultrafast internal conversion, this state converts to the lower-lying state S₁ with 0.7 eV of vibrational energy. The expected ionization propensity rules, according to Fig. 1, are shown: S₂ $\rightarrow$ D₀ + e⁻ (ϵ_1) and S₁ $\rightarrow$ D₁ + e⁻ (ϵ_2). The dashed lines indicate electron emission. **b**, Femtosecond time-resolved photoelectron kinetic energy (E_{kin}) spectra of DT pumped at 287 nm and probed at 235 nm. There is a rapid shift (~ 400 fs) in the distribution: from (ϵ_1), an energetic peak at 2.5 eV due to photoionization of S₂ into the D₀ cation ground electronic state; to (ϵ_2), a broad, structured band at lower energies due to photoionization of vibrationally hot S₁ into the D₁ cation first excited electronic state. The structure in the low-energy band reflects the vibrational dynamics in S₁. The time delay between the pump (excitation) and the probe (ionization) laser pulses is indicated by Δt .

We note that the results in Fig. 2 contain much more information than the overall (integrated) internal conversion time. It can be seen that there is vibrational structure in the photoelectron spectra. This structure yields information about the state-to-state vibrational dynamics which promotes and tunes the electronic population transfer, and the ensuing intramolecular vibrational energy redistribution in the 'hot' molecule which occurs on the S₁ potential surface (subjects of our current theoretical investigations).

We have confirmed the above results by comparing them with another scheme using both one- and two-photon ionization. In this case, the first probe photon is above the ionization potential but below D₁ (Fig. 3a). As above, the S₂ state (again prepared at 287 nm) ionizes into D₀, here producing the slow, sharp electron band seen at $\epsilon_1 = 0.4$ eV in Fig. 3b. This band is the same as the $\epsilon_1 = 2.5$ eV band from Fig. 2, but is shifted to lower energy due to the lower-energy probe photon. The band decays rapidly with a time constant of 377 ± 47 fs, providing an independent determination of the S₂ $\rightarrow$ S₁ conversion time in agreement with that of Fig. 2. The D₁ ion state is no longer energetically accessible from S₁ via a single photon. At 352 nm, however, our probe laser intensity was sufficient to allow absorption of a second photon. Two-photon ionization accesses different ion electronic states due to the symmetry of the two-

photon electric dipole operator, as indicated in Fig. 3a, producing energetic electrons. The S_1 correlations upon two-photon probe ionization can now also include D_0 and D_2 (a higher-lying ion state), producing the broad, energetic ϵ_2 electron bands in Fig. 3b. The switching (at invariant laser intensity) from one-photon to a two-photon favoured ionization again indicates the selectivity of the photoionization process for given electronic state symmetries.

Many photoinduced polyatomic unimolecular reactions are based on vibrationally 'hot' molecules formed via ultrafast internal conversion²⁶. A fundamental question in unimolecular reaction rate theory is that of the assumption of statistical energy redistribution due to intramolecular vibrational energy redistribution being very fast compared with reaction²⁷. Many models and interpretations are based upon the supposition of fast, complete intramolecular vibrational energy redistribution. Increasingly, however, this assumption has come into question²⁸. We believe that our method could present new views of the extent and timescale for the onset of statistical behaviour in an isolated, energized molecule.

Finally, we note possible extensions of the technique we report here. If, for symmetry reasons, the coupled neutral electronic configurations correlate to the same ion electronic configuration, then the electronic–nuclear mixing cannot be clearly discerned by a

photoelectron kinetic-energy analysis alone. In such cases, the photoelectron angular (that is, partial wave) distributions^{29,30}, which reflect electron angular-momentum correlations on ionization, should in general still change during a non-adiabatic process. The molecular ionization continuum could provide an interesting template for the study of time-resolved molecular dynamics: ionic vibrational and rotational structure are sensitive to nuclear dynamics, while the ion electronic and the electron partial wave structure are sensitive to electronic population dynamics. We hope that these experimental methods, when combined with dynamical calculations, will provide new insights into the non-adiabatic processes that underpin our understanding of photochemistry, material photosciences and molecular electronics. □

Received 10 September 1998; accepted 6 July 1999.

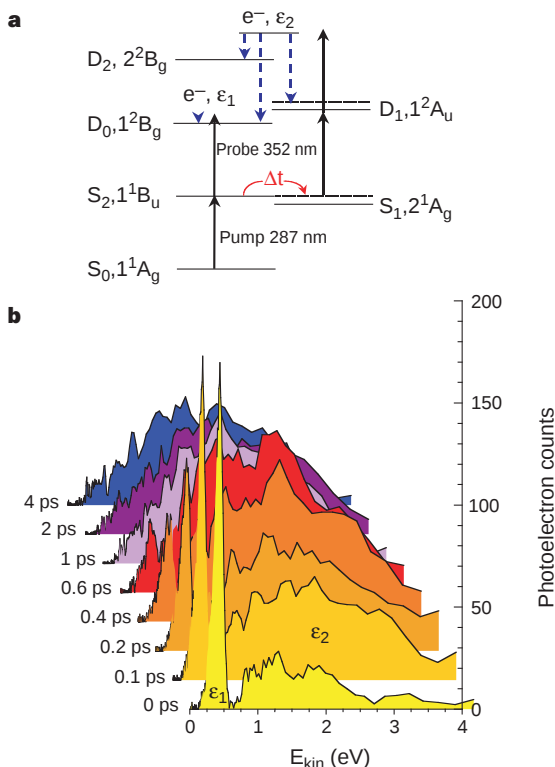


Figure 3 Time-resolved vibrational and electronic dynamics in DT using two-photon ionization. **a**, Level scheme in DT for one- and two-photon probe ionization. The pump laser is identical to that in Fig. 2 and prepares the same S_2 state wavepacket. The expected ionization propensity rules are: $S_2 \rightarrow D_0 + e^- (\epsilon_1)$ for one-photon ionization (as in Fig. 2) and $S_1 \rightarrow D_0, D_1, D_2 + e^- (\epsilon_2)$ for two-photon ionization. The dashed lines indicate electron emission. **b**, Femtosecond time-resolved photoelectron kinetic energy spectra of DT pumped at 287 nm and probed at 352 nm, using both one- and 2-photon probes. At 352 nm, the D_1 ion state is not energetically accessible from the S_1 state via a single photon transition. Confirming the results of Fig. 2, there is a rapid shift (~ 400 fs) in the distribution: from (ϵ_1) a peak at 0.4 eV due to one-photon ionization of S_2 into the D_0 cation ground electronic state; to (ϵ_2) a broad, structured band at higher energies (1–3.5 eV) due to two-photon ionization of the vibrationally hot S_1 state into the D_0 cation ground and excited electronic states. The photoionization channel switches from a one-photon to a two-photon process during the internal conversion, indicating again that the electronic structure of the ionization continuum is selective of the evolving electronic symmetry in the neutral state.

- Rischel, C. *et al.* Femtosecond time-resolved X-ray diffraction from laser-heated organic films. *Nature* **390**, 490–492 (1997).
- Williamson, J. C., Jianming, C., Hyotcherl, I., Frey, H. & Zewail, A. H. Clocking transient chemical changes by ultrafast electron diffraction. *Nature* **386**, 159–162 (1997).
- Stepfeldt, H., Constant, E. & Corkum, P. B. Wavepacket structure and dynamics measured by Coulomb explosion. *Phys. Rev. Lett.* **74**, 3780–3783 (1995).
- Jortner, J., Rice, S. A. & Hochstrasser, R. M. Radiationless transitions in photochemistry. *Adv. Photochem.* **7**, 149–309 (1969).
- Stock, G. & Domcke, W. Theory of ultrafast nonadiabatic excited state processes and their spectroscopic detection in real time. *Adv. Chem. Phys.* **100**, 1–169 (1997).
- Michl, J. & Bonacic-Koutecky, J. *Electronic Aspects of Organic Photochemistry* (Wiley, New York, 1990).
- Schoenlein, R. W., Peteanu, L. A., Mathies, R. A. & Shank, C. V. The first step in vision: femtosecond isomerization of rhodopsin. *Science* **254**, 412–415 (1991).
- Jortner, J. & Ratner, M. A. *Molecular Electronics* (IUPAC, Blackwell, Oxford, 1997).
- Zewail, A. H. *Femtochemistry: Ultrafast Dynamics of the Chemical Bond* (World Scientific, Singapore, 1994).
- Fischer, I., Villeneuve, D. M., Vrakking, M. J. J. & Stolow, A. Wavepacket dynamics studied by time-resolved zero-kinetic energy photoelectron spectroscopy. *J. Chem. Phys.* **102**, 5566–5569 (1995).
- Blanchet, V. & Stolow, A. Non-adiabatic dynamics in polyatomic systems studied by femtosecond time-resolved photoelectron spectroscopy. *J. Chem. Phys.* **108**, 4371–4374 (1998).
- Seel, M. & Domcke, W. Femtosecond time-resolved ionization spectroscopy of ultrafast internal-conversion dynamics in polyatomic molecules: theory and computational studies. *J. Chem. Phys.* **95**, 7806–7822 (1991).
- Cyr, D. R. & Hayden, C. C. Femtosecond time-resolved photoionization and photoelectron spectroscopy studies of ultrafast internal conversion in 1,3,5-hexatriene. *J. Chem. Phys.* **104**, 771–774 (1996).
- Radloff, W. *et al.* Internal conversion in highly excited benzene and benzene dimer: femtosecond time-resolved photoelectron spectroscopy. *Chem. Phys. Lett.* **281**, 20–26 (1997).
- Kim, B., Schick, C. P. & Weber, P. M. Time-delayed two-colour photoelectron spectra of aniline, 2-aminopyridine and 3-aminopyridine: Snapshots of the nonadiabatic curve crossing. *J. Chem. Phys.* **103**, 6903–6913 (1995).
- Baumert, T., Thalweiser, R. & Gerber, G. Femtosecond two-photon ionization spectroscopy of the B-state of Na_3 clusters. *Chem. Phys. Lett.* **209**, 29–34 (1993).
- Assion, A. *et al.* Femtosecond pump-probe photoelectron spectroscopy: Mapping of vibrational wavepacket motion. *Phys. Rev. A* **54**, R4605–R4608 (1996).
- Hudson, B. S., Kohler, B. E. & Schulten, K. *Excited States Vol. 6, Linear Polyene Electronic Structure and Potential Surfaces* (ed. Lim, E. C.) (Academic, New York, 1982).
- Orlandi, G., Zerbetto, F. & Zgierski, M. Z. Theoretical analysis of spectra of short polyenes. *Chem. Rev.* **91**, 867–891 (1991).
- Petek, H., Bell, A. J. & Yoshihara, K. & Christensen, R. L. Spectroscopic and dynamical studies of the S_1 and S_2 states of decatetraene in supersonic molecular beams. *J. Chem. Phys.* **95**, 4739–4749 (1991).
- Warshel, A. & Karplus, M. Calculations of ground and excited state potential surfaces of conjugated molecules. I. Formulation and parametrization. *J. Am. Chem. Soc.* **94**, 5612–5625 (1972).
- Zerbetto, F., Zgierski, M. Z., Negri, F. & Orlandi, G. Theoretical study of the force fields of the lowest three singlet electronic states of linear polyenes. *J. Chem. Phys.* **89**, 3681–3688 (1988).
- Andrews, J. R. & Hudson, B. S. Polyene spectroscopy: vibronic evidence for a forbidden transition in deca-2,4,6,8-tetraene. *Chem. Phys. Lett.* **57**, 600–604 (1979).
- Fronzoni, G., Decleva, P., Lisini, A. & De Alti, G. Ab initio CI investigation of valence photoelectron spectra of conjugated hydrocarbons. *J. Electron. Spectrosc. Relat. Phenom.* **69**, 207–224 (1994).
- Bouwman, W. G. *et al.* Fluorescence of gaseous tetraenes and pentaenes. *J. Phys. Chem.* **94**, 7429–7434 (1990).
- Wild, W., Seilmeier, A., Gottfried, N. H. & Kaiser, W. Ultrafast investigations of vibrationally hot molecules after internal conversion in solution. *Chem. Phys. Lett.* **119**, 259–263 (1985).
- Forst, W. *Theory of Unimolecular Reactions* (Academic, New York, 1973).
- Remacle, F. & Levine, R. D. Does a dissociating molecule sample the available phase space? *J. Phys. Chem.* **95**, 7124–7127 (1991).
- Reid, K. L., Duxon, S. P. & Towrie, M. Observation of time- and angle-resolved photoelectron flux from an optically prepared state of a molecule: hyperfine depolarization in $NO A^1\Sigma^+$. *Chem. Phys. Lett.* **228**, 351–356 (1994).
- Seideman, T. Time-resolved photoelectron angular distributions: a nonperturbative theory. *J. Phys. Chem.* **107**, 7859–7868 (1997).

Acknowledgements

We thank P. Arya and M. Barnes for the synthesis of all-*trans* 2,4,6,8 decatetraene. A.S. thanks W. Domcke, P. B. Corkum and M. Yu. Ivanov for discussions.

Correspondence and requests for materials should be addressed to A.S. (e-mail: albert.stolow@nrc.ca).

Small-bandgap endohedral metallofullerenes in high yield and purity

S. Stevenson*, G. Rice*, T. Glass*, K. Harich*, F. Cromer*, M. R. Jordan*, J. Craft*, E. Hadju†, R. Bible†, M. M. Olmstead‡, K. Maitra‡, A. J. Fisher‡, A. L. Balch‡ & H. C. Dorn*

* Department of Chemistry, Virginia Polytechnic Institute of State University, Blacksburg, Virginia 24061, USA

† G. D. Searle & Co., 5200 Old Orchard Road, Skokie, Illinois 60077, USA

‡ Department of Chemistry, University of California, Davis, California 95616, USA

The idea¹ that fullerenes might be able to encapsulate atoms and molecules has been verified by the successful synthesis of a range of endohedral fullerenes, in which metallic or non-metallic species are trapped inside the carbon cage^{2–13}. Metal-containing endohedral fullerenes have attracted particular interest as they might exhibit unusual material properties associated with charge transfer from the metal to the carbon shell. However, current synthesis methods have typical yields of less than 0.5%, and produce multiple endohedral fullerene isomers, which makes it difficult to perform detailed studies of their properties. Here we show that the introduction of small amounts of nitrogen into an electric-arc reactor allows for the efficient production of a new family of stable endohedral fullerenes encapsulating trimetallic nitride clusters, $\text{Er}_x\text{Sc}_{3-x}\text{N@C}_{80}$ ($x = 0–3$). This ‘trimetallic nitride template’ process generates milligram quantities of product containing 3–5% $\text{Sc}_3\text{N@C}_{80}$, which allows us to isolate the material and determine its crystal structure, and its optical and electronic properties. We find that the Sc_3N moiety is encapsulated in a highly symmetric, icosahedral C_{80} cage, which is stabilized as a result of charge transfer between the nitride cluster and the fullerene cage. We expect that our method will provide access to a range of small-bandgap fullerene materials, whose electronic properties can be tuned by encapsulating nitride clusters containing different metals and metal mixtures.

We have previously noted an unidentified scandium endohedral metallofullerene species (with a mass-to-charge ratio, m/z , of 1,109) in chromatographic fractions that were obtained from the extraction of scandium metallofullerene soot¹⁴. We now report that introduction of a small amount of N_2 gas into the Krätschmer–Huffman generator¹⁵ results in significantly enhanced product of this species. The yields even exceed those of the usually most abundant species, the empty-cage C_{84} , with a single packed-graphite rod, the soluble fullerene and endohedral metallofullerene fraction (~60 mg) results in the isolation of 2–4 mg of the species with $m/z = 1109$ (see Methods for details on synthesis and isolation). The isotope patterns for the negative-ion mass spectra for natural abundance (Fig. 1c) and ^{13}C -enriched samples of this species are consistent with $\text{Sc}_3\text{N@C}_{80}$.

The structures of the empty icosahedral C_{80} cage and of $\text{Sc}_3\text{N@C}_{80}$ are shown in Fig. 1a and b, respectively, while the 150 MHz ^{13}C NMR spectrum for the ~11% ^{13}C enriched $\text{Sc}_3\text{N@C}_{80}$ sample is shown in Fig. 1d. As predicted for a C_{80} cage with I_h symmetry⁸, the spectrum consists of only two resolved lines (at chemical shifts of 144.57 and 137.24 p.p.m.) with an intensity ratio of 3/1. This ^{13}C NMR spectrum contrasts with the single line reported by Akasaka *et al.*¹³ for $\text{La}_2\text{@C}_{80}$; they interpreted their result in terms of rapid circular motion of the La atoms in a C_{80} cage with I_h symmetry. Fullerene carbon atoms at the intersection of three six-membered rings (6MR), pyrene-type sites (shown in dark blue in Fig. 1a), usually exhibit ^{13}C chemical shifts in the range 130–138 p.p.m. (ref. 16) which is consistent with the peak at 137.24 p.p.m. Carbon atoms

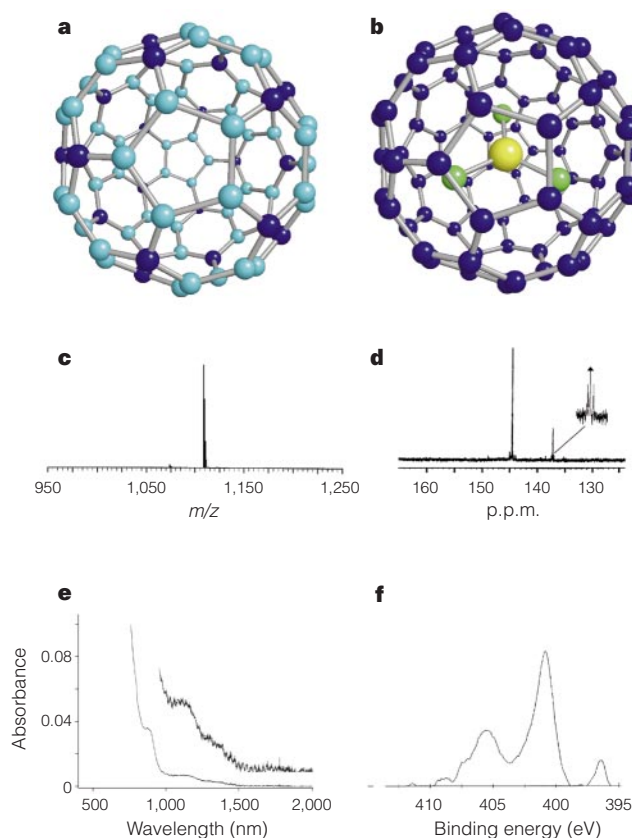


Figure 1 Structures and spectroscopic data for $\text{Sc}_3\text{N@C}_{80}$. **a**, Empty-cage fullerene, C_{80} , I_h isomer, showing pyrene (dark blue) and corannulene (light blue) sites; **b**, structure of $\text{Sc}_3\text{N@C}_{80}$. **c**, Negative ion-desorption chemical ionization (NI–DCI) mass spectrum for purified $\text{Sc}_3\text{N@C}_{80}$ (see Supplementary Information for mass spectral isotope data and HPLC purity trace). **d**, The 150 MHz ^{13}C NMR spectrum (17,130 scans, ~15 hours) for purified, ^{13}C -labelled (11%) $\text{Sc}_3\text{N@C}_{80}$ in CS_2/CrIII acetylacetonate and referenced to tetramethyl silane (TMS). **e**, Ultraviolet-visible-near infrared spectrum for $\text{Sc}_3\text{N@C}_{80}$ in CS_2 ; inset, vertical expansion. **f**, X-ray photoelectron spectroscopy (XPS) spectrum of $\text{Sc}_3\text{N@C}_{80}$ evaporated on a Au film.

at the intersection of one five-membered ring (5MR) and two 6MRs, corannulene sites (shown in light blue in Fig. 1a), are usually found in the range 138–145 p.p.m. which is consistent with the peak at 144.57 p.p.m. As expected, the observed ^{13}C – ^{13}C spin–spin coupling pattern ($^1J_{\text{C–C}} = 57.2 \text{ Hz}$) has the same satellite intensity for the doublets centred about the central peaks at 144.57 and 137.24 p.p.m. (3/1 intensity ratio). This result is consistent with a pyrene site (137.24 p.p.m.) bonded to three equivalent corannulene sites (Fig. 1a). The corannulene site (144.57 p.p.m.) exhibits coupling to only one adjacent pyrene site. The magnitude of this scalar coupling interaction (57.2 Hz) is similar to values reported for other fullerene sp^2 – sp^2 hybridized carbons¹⁷. Our ^{13}C NMR results suggest that internal motion of the Sc_3N cluster yields a time-averaged electronic environment that preserves overall I_h symmetry for the carbon cage, with the encapsulated Sc_3N cluster not localized at specific internal bonding sites (on the NMR timescale at 295 K). An ambient temperature, the ^{45}Sc NMR spectrum for $\text{Sc}_3\text{N@C}_{80}$ exhibits a single symmetric line which is also indicative of a dynamic structure (see Supplementary Information). A unique feature of the C_{80} cage, with I_h symmetry, is the absence of bonding between two different 5MRs as illustrated by the isolated 5MRs (light blue) in Fig. 1a. These bonded sites between 5MRs are usually the most reactive in common fullerene cages (C_{60} , C_{70} and C_{84}) and a different chemical reactivity pathway will be necessary for $\text{Sc}_3\text{N@C}_{80}$ (ref. 18). Almost all fullerenes react with inorganic and organic reagents by addition to 6:6 ring junctions between two pentagons. These are absent in $I_h\text{C}_{80}$.

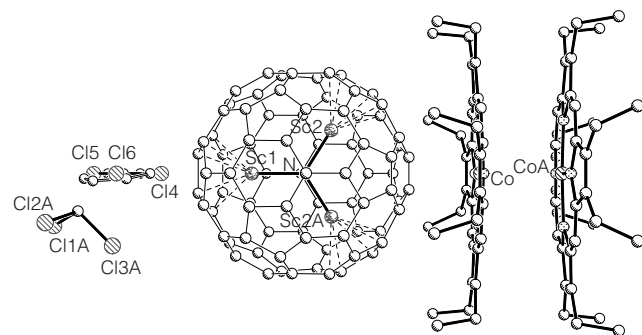


Figure 2 The structure of $(\text{Sc}_3\text{N}@\text{C}_{80})\cdot\text{Co}^{\text{II}}(\text{OEP})\cdot 1.5 \text{CHCl}_3\cdot 0.5 \text{C}_6\text{H}_6$, perpendicular to the crystallographic mirror plane that bisects Co, Sc1 and N. The porphyrin portion has normal distances and angles. Only the main orientation (0.75 fractional occupancy) of the Sc_3N portion is shown. Bond distances ($\text{\AA}$): N–Sc1, 2.011(19); N–Sc2, 1.966(12); shortest C–Sc, 2.188(9), 2.170(10). Angles (deg): Sc1–N–Sc2, 122.2(6); Sc2–N–Sc2A, 115.5(11). Benzene and chloroform (Cl4–Cl6) molecules are randomly distributed over a common site in a 1:3 ratio.

The central electronegative nitrogen atom in the Sc_3N cluster significantly alters other properties of this new family of endohedral metallofullerenes. It has been established^{19,20} that the chromatographic retention parameter k ($= t_r/t_0$ = retention time/void time) is proportional to the polarizability and number of π -electrons at the fullerene cage surface for non-polar chromatographic stationary phases such as 2-(1-pyrenyl)ethyl (PYE), pyrenylethyl, and pentabromobenzyl (PBB)^{19,20}. $\text{La}_2@\text{C}_{80}$ on a PBB stationary phase (toluene solvent) has, for example, a reported²¹ elution time corresponding to that observed for empty-cage $\text{C}_{88}\text{--}\text{C}_{90}$ whereas $\text{Sc}_3\text{N}@\text{C}_{80}$ elutes together with $\text{C}_{84}\text{--}\text{C}_{86}$ under the same conditions. The different behaviour of $\text{Sc}_3\text{N}@\text{C}_{80}$ corresponds to a decrease of approximately four π -electrons at the fullerene cage surface relative to $\text{La}_2@\text{C}_{80}$ and suggests significant electron withdrawal from the carbon cage surface by the central nitrogen atom.

Although the electronic spectra of most endohedral metallofullerenes are dominated by features of the carbon cage rather than the encapsulated metals, the spectrum of the Sc_3N cluster indicates significant perturbation of the carbon cage. The ultraviolet-visible-near infrared spectrum of $\text{Sc}_3\text{N}@\text{C}_{80}$ (Fig. 1e) consists of significant absorption beyond 1,000 nm, and prominent peaks at 900 and 1,140 nm. If the absorption onset corresponds to the bandgap energy, the onset for $\text{Sc}_3\text{N}@\text{C}_{80}$ (at $\sim 1,560$ nm) corresponds to a bandgap of only 0.8 eV. The absorption onset for $\text{La}_2@\text{C}_{80}$ is significantly higher in energy ($\sim 1,000$ nm, 1.3 eV), and is consistent with the larger bandgap reported for this endohedral species^{21,22}. The X-ray photoelectron spectroscopy (XPS) spectrum for $\text{Sc}_3\text{N}@\text{C}_{80}$ (Fig. 1f) has an absorption at ~ 400.9 eV for the $2p_{3/2}\text{Sc}$ core level, and a second peak (due to spin–orbital coupling) 4.8 eV higher in energy that is consistent with other scandium endohedral species. A small nitrogen peak is observed at 396.4 eV. The scandium and nitrogen XPS signal areas (corrected for relative sensitivities) indicate a 3/1 Sc/N ratio for the cluster. In addition, the binding energy for the nitrogen peak (396.4 eV) compares favourably with the value reported for scandium nitride²³, 396.2 eV.

We confirmed the endohedral nature of $\text{Sc}_3\text{N}@\text{C}_{80}$ by a single-crystal X-ray diffraction study of $(\text{Sc}_3\text{N}@\text{C}_{80})\cdot\text{Co}^{\text{II}}(\text{OEP})\cdot 1.5 \text{CHCl}_3\cdot 0.5 \text{C}_6\text{H}_6$. Discrepancies between observed and calculated data remain, but appear to be accounted for by the presence of different orientations of the C_{80} cage. (See Methods and Supplementary Information for details). Figure 2 shows the structure, with the planar Sc_3N unit clearly encapsulated within the C_{80} cage. The $\text{Sc}_3\text{N}@\text{C}_{80}$ molecule is close to (but not covalently bound to) the $\text{Co}^{\text{II}}(\text{OEP})$ molecule, which makes face-to-face contact with another $\text{Co}^{\text{II}}(\text{OEP})$ moiety, as seen in related compounds²⁴. The N–Sc distances, 2.011 (± 0.019), 1.966 (± 0.012) $\text{\AA}$, are slightly

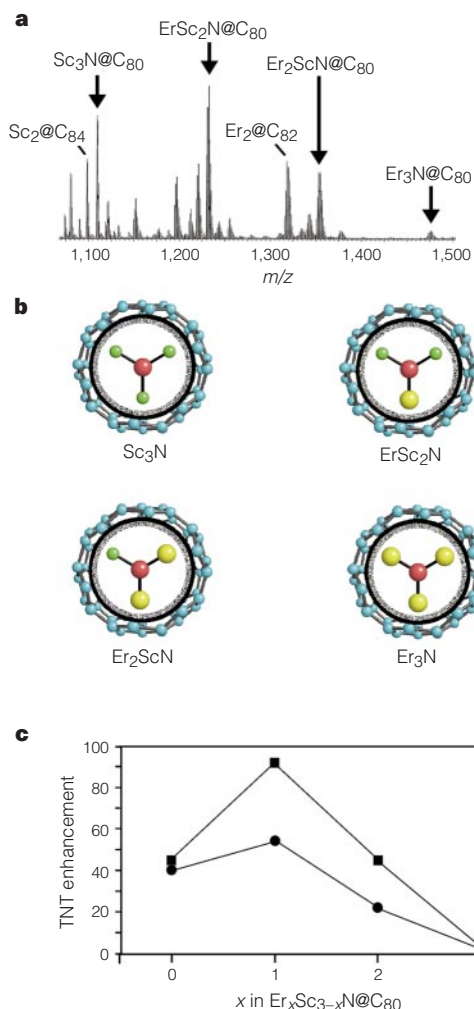


Figure 3 Mass spectral data for $\text{Er}_{3-x}\text{Sc}_x\text{N}@\text{C}_{80}$ formation. **a**, The NI-DCl mass spectrum for the mixed scandium and erbium endohedral metallofullerene chromatographic fraction after the first automated PBB/carbon disulphide separation stage. This fraction was maximized to illustrate all species (see Supplementary Information) for mass spectral isotope data. **b**, The pictorial cut-out representation of $\text{Sc}_3\text{N}@\text{C}_{80}$, $\text{ErSc}_2\text{N}@\text{C}_{80}$, $\text{Er}_2\text{ScN}@\text{C}_{80}$ and $\text{Er}_3\text{N}@\text{C}_{80}$. **c**, Plot showing the higher yields obtained with our 'trimetallic nitride template' (TNT) process. Data are shown for graphite rods packed with two different metal loadings, distinguished by their $\text{Er}_2\text{O}_3/\text{Sc}_2\text{O}_3$ /powdered graphite ratio: filled circles, 1.5/1.5/97; filled squares, 3/3/94. The yields obtained from the mass spectral data are internally referenced to the well known trimetallic endohedral¹⁴, $\text{Sc}_3@\text{C}_{82}$, which is also formed in the absence of nitrogen (not formed via the TNT process). Yields were determined from NI-DCl mass spectrometry.

shorter than the Sc–N bonds in amide compounds such as $\{(\text{HSiMe}_2)_2\text{N}\}_3\text{Sc}(\text{THF})$ (average Sc–N distance, 2.069 $\text{\AA}$)²⁵. In the solid, the scandium ions face three pentagons within the C_{80} cage.

Graphite rods were also packed with an $\text{Er}_2\text{O}_3/\text{Sc}_2\text{O}_3$ /powered graphite mixture, and vaporized in similar fashion in the Krätschmer–Huffman generator. In Fig. 3a we show the negative ion-desorption chemical ionization mass spectrum for the chromatographic fraction after the first automated PBB/carbon disulphide separation stage (see Methods) which corresponds to elution times for empty-cage $\text{C}_{82}\text{--}\text{C}_{86}$ species. All members of the $\text{Er}_x\text{Sc}_{3-x}\text{N}@\text{C}_{80}$ ($x = 0\text{--}3$) family are present in this fraction (see also Fig. 3b). The higher yields with our 'trimetallic nitride template' process are illustrated in Fig. 3c for graphite rods packed with two different metal loadings. $\text{Sc}_3\text{N}@\text{C}_{80}$, $\text{Er}_2\text{ScN}@\text{C}_{80}$ and $\text{ErSc}_2\text{N}@\text{C}_{80}$ are all formed in yields that are 1–2 orders of magnitude greater than $\text{Sc}_3@\text{C}_{82}$. Besides the previously described $\text{Er}_2@\text{C}_{82}$ and $\text{Sc}_2@\text{C}_{84}$

species^{3,26}, the mixed metal endohedrals ErSc@C₈₂ and ErSc@C₈₄ also appear to be members of this fraction, but at nearly the same levels as the usually prominent endohedral, Sc₂@C₈₄. □

Methods

Synthesis and isolation.

Graphite rods filled with Sc₂O₃/graphite (3 wt%/97 wt%) were vaporized in a Krätschmer–Huffman generator¹⁵ in a dynamic helium atmosphere (300 torr) containing a small amount of nitrogen (1–3 torr). A single rod typically yields 60 mg soluble fullerenes and endohedral metallofullerenes, which are extracted from the soots using carbon disulphide. Following extraction of the fullerene and metallofullerene containing soots, the ¹³C-labelled and natural-abundance extracts were filtered through glass wool. The initial automated chromatographic separation stage utilized a pentabromobenzyl column (PBB, 25 cm × 10 mm inside diameter; Phenomenex Co., Torrance, California, USA) with CS₂ as the mobile phase (2 ml min⁻¹). In the second and third stages, a selective semi-preparative Trident-Tri-DNP (di-nitrophenyl) column (Buckyclutcher, 25 cm × 10 mm inside diameter, Regis Chemical, Morton Grove, Illinois, USA) was utilized with toluene as the solvent (2 ml min⁻¹). The final purification stage involved separation with the PBB column (CS₂ mobile phase, 2 ml min⁻¹). Carbon disulphide solutions of Sc₃N@C₈₀ (solubility, ~2 mg ml⁻¹) are reddish-brown. Sc₃N@C₈₀ is soluble in other non-polar solvents (toluene, decalin), but has no significant solubility in more polar solvents (diethyl ether, acetone, methanol).

X-ray analysis.

The X-ray diffraction study was performed on (Sc₃N@C₈₀)-Co^{II}(OEP)-1.5 CHCl₃·0.5 C₆H₆, which was obtained by mixing a solution of Sc₃N@C₈₀ in benzene/carbon disulphide with a solution of Co^{II}(OEP) (OEP is the dianion of octaethylporphyrin) in chloroform. Crystal data as follows. fw = 1,919.48, dark red parallelepiped, 0.02 × 0.02 × 0.02 mm, monoclinic, space group C2/m, *a* = 25.142(5) Å, *b* = 15.246(3) Å, *c* = 19.459(3) Å, β = 94.79(3)°, *V* = 7433(3) Å³, λ = 1.54178 Å, *Z* = 4, *D*_c = 1.715 Mg m⁻³; μ(Cu-Kα) = 6.012 mm⁻¹; 2θ_{max} = 113°; *T* = 130 K; 9,878 refl. collected; 5,138 independent (*R*_{int} = 0.112) included in the refinement; no absorption correction performed; programs used for solution and refinement, SHELXS-97 (ref. 27); full-matrix least squares based on *F*², SHELXL-97 (ref. 27); 312 parameters, *R*₁ = 0.2730, *wR*₂ = 0.5932 for all data; *R*₁ = 0.2244 computed for 3,781 observed data (>2σ(*I*)). The *R* values did not fall to satisfactory levels. Difference Fourier maps suggest that the discrepancy resides in electron density in the region of the C₈₀ cage, and that other orientations of the cage are also present (see Supplementary Information for further details).

Received 22 March; accepted 19 July 1999.

- Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. Buckminsterfullerene. *Nature* **318**, 162–163 (1985).
- Chai, Y. *et al.* Fullerenes with metals inside. *J. Phys. Chem.* **95**, 7564–7560 (1991).
- Beyers, R. *et al.* Preparation and structure of the metallofullerene Sc₂@C₈₄. *Nature* **370**, 196–199 (1994).
- Bethune, D. S., Johnson, R. D., Salem, J. R., de Vries, M. S. & Yannoni, C. S. Atoms in carbon cages: the structure and properties of endohedral fullerenes. *Nature* **366**, 123–128 (1993).
- Nagase, S., Kobayashi, K. & Akasaka, T. in *Fullerenes: Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials* (eds Kadish, K. M. & Ruoff, R. S.) 747–762 (Electrochemical Society, Pennington, 1995).
- Nagase, S., Kobayashi, K. & Akasaka, T. Endohedral metallofullerenes: new spherical cage molecules with interesting properties. *Bull. Chem. Soc. Jpn* **69**, 2131–2142 (1996).
- Manolopoulos, D. E. & Fowler, P. W. Structural proposals for endohedral metal-fullerene complexes. *Chem. Phys. Lett.* **187**, 17 (1991).
- Fowler, P. W. & Manolopoulos, D. E. An Atlas of Fullerenes (Oxford Univ. Press, 1995).
- Heinrich, F. H. *et al.* Isolation and characterization of C₈₀. *Angew. Chem. Int. Edn Engl* **35**, 1732–1734 (1996).
- Wang, C.-R., Dennis, J. S., Inakura, M. & Shinohara, H. in *Fullerenes: Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials* (eds Kadish, K. M. & Ruoff, R. S.) 1023–1030 (Electrochemical Society, Pennington, 1998).
- Kobayashi, K. & Nagase, S. Structures and electronic states of endohedral dimetallofullerenes: M₂@C₈₀ (M = Sc, Y, La, Ce, Pr, Eu, Gd, Yb, and Lu). *Chem. Phys. Lett.* **262**, 227–232 (1996).
- Kobayashi, K., Nagase, S. & Akasaka, T. Endohedral dimetallofullerenes Sc₂@C₈₄ and La₂@C₈₀. Are the metal atoms still inside the fullerene cages? *Chem. Phys. Lett.* **261**, 502–506 (1996).
- Akasaka, T. *et al.* ¹³C and ¹³⁹La NMR studies of La₂@C₈₀: first evidence for circular motion of metal atoms in endohedral metallofullerenes. *Angew. Chem. Int. Edn Engl* **36**, 1643–1645 (1997).
- Stevenson, S. *et al.* Automated HPLC separation of endohedral metallofullerene Sc@C_{2n} and Y@C_{2n} fractions. *Anal. Chem.* **66**, 2675–2679 (1994).
- Krätschmer, W., Fostiropoulos, K. & Huffman, D. R. The infrared and ultraviolet absorption spectra of laboratory-produced carbon dust: evidence for the presence of the C₆₀ molecule. *Chem. Phys. Lett.* **170**, 167–170 (1990).
- Diederich, F. & Whetten, R. L. Beyond C₆₀: the higher fullerenes. *Acc. Chem. Res.* **25**, 119–126 (1992).
- Johnson, R. D., Meier, G., Salem, J. R. & Bethune, D. S. 2D nuclear magnetic resonance study of the structure of the fullerene C₇₀. *J. Am. Chem. Soc.* **113**, 3619–3621 (1991).
- Hirsch, A. *The Chemistry of the Fullerenes* (Thieme Verlag, Stuttgart, 1994).
- Fuchs, D. *et al.* Extraction and chromatographic elution behavior of endohedral metallofullerenes: inferences regarding effective dipole moments. *J. Phys. Chem.* **100**, 725–729 (1996).
- Stevenson, S. *et al.* La₂@C₇₂: metal-mediated stabilization of a carbon cage. *J. Phys. Chem.* **102**, 2833–2837 (1998).
- Yamamoto, K., Ishiguro, T., Sakurai, K. & Funasaka, H. in *Fullerenes: Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials* (eds Kadish, K. M. & Ruoff, R. S.) 743–753 (Electrochemical Society, Pennington, 1997).
- Kikuchi, K., Nakao, Y., Achiba, Y. & Nomura, M. in *Fullerenes: Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials* (eds Kadish, K. M. & Ruoff, R. S.) 1300–1308 (Electro-

chemical Society, Pennington, 1994).

- Wagner, C. D., Riggs, W. M., Davis, L. E., Moulder, J. F. & Muilenberg, G. E. (eds) *Handbook of X-ray Photoelectron Spectroscopy* (Perkin-Elmer Corporation, Eden Prairie, Minnesota, 1978).
- Olmstead, M. M. *et al.* Interaction of curved and flat molecular surfaces. The structures of crystalline compounds composed of fullerene (C₆₀, C₆₀O, C₇₀, and C₁₂₀O) and metal octaethylporphyrin units. *J. Am. Chem. Soc.* **121**, 7090–7097 (1999).
- Anwander, R. *et al.* Synthesis and structural characterization of rare-earth bis(dimethylsilyl)amides and their surface organometallic chemistry on mesoporous MCM-41. *J. Chem. Soc. Dalton Trans.* 847–858 (1998).
- Dorn, H. C. *et al.* Endohedral metallofullerenes: isolation and characterization. *Proc. Mater. Res. Soc.* **359**, 123–135 (1995).
- Sheldrick, G. M. Phase Annealing in SHELX: direct methods for larger structures. *Acta Crystallogr. A* **46**, 467–473 (1990).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank P. Burbank and Z. Sun for technical support. H.C.D. acknowledges discussions with M. Sherwood, D. S. Bethune, M. Anderson and R. Heflin, and thanks the NSF and the Virginia Tech ASPIRES program for supporting initial phases of this study. A.L.B. thanks the NSF for support.

Correspondence and requests for materials should be addressed to H.C.D. (e-mail: hdorn@chemserver.chem.vt.edu).

A Middle Jurassic mammal from Madagascar

John J. Flynn*, J. Michael Parrish†, Berthe Rakotosamimanana‡, William F. Simpson* & André R. Wyss§

* Department of Geology, Field Museum of Natural History, Chicago, Illinois 60605, USA

† Department of Biological Sciences, Northern Illinois University, DeKalb, Illinois 60115, USA

‡ Département de Paléontologie et Anthropologie Biologique, Université d'Antananarivo, Antananarivo 101, Madagascar

§ Department of Geological Sciences, University of California, Santa Barbara, California 93106, USA

The lower molars of tribosphenic mammals (marsupials, placentals and their extinct allies) are marked, primitively, by a basined heel (talonid) acting as the mortar to the pestle of a large inner cusp (protocone) on the opposing upper teeth. Here we report the earliest tribosphenic mammal found so far, three lower teeth in a jaw fragment from Middle Jurassic (Bathonian, ~167 ± 2 Myr)¹ sediments of northwest Madagascar. This specimen extends the stratigraphic range of the Tribosphenida by some 25 million years, more than doubling the age of the oldest mammal known from Madagascar², and representing only the second pre-Plio/Pleistocene terrestrial mammal known from the island. Although it indicates a more ancient diversification of the Tribosphenida than previously thought, this find fails to confirm molecular-clock-based models proposing a Middle Jurassic divergence of marsupials and placentals³. In addition, it offers a glimpse of mammal evolution on the southern continents during the Middle through Late Jurassic, countering the prevailing view⁴ of a northern origin for tribosphenic mammals.

The miniscule mammal reported here is from the youngest of a series of early Mesozoic vertebrate faunas discovered in the Isalo 'Group' of western Madagascar⁵. The locality is in the upper level of the Isalo's tripartite division (Isalo III) in the Mahajanga Basin. It lies near the top of the 'Bathonian Facies Mixte Dinosauriens' of Isalo IIb, the age of which is based on a rich invertebrate fauna⁶ including the diagnostic echinoids *Nucleolites amplius* and *Acrosalenia colcanapi*. 'Crocodile' and plesiosaur teeth, plus fragmentary remains of the sauropod dinosaur *Lapparentosaurus*

(= *Bothriospondylus*)⁷, are the only vertebrates previously reported from the Isalo III—a unit exposed over a several-hundred-kilometre-long arc along the eastern and southern margins of the basin.

The closest approximation to the basined, cusp-ringed talonids of tribosphenidans are the expanded but lingually open structures in *Palaeoxonodon*, *Peramus* and *Arguimus*⁸ (Fig. 1). Discovery of the Isalo IIIb jaw, which possesses some but not all of the dental specializations usually ascribed to tribosphenidans, calls attention to the previous lack of a formal definition for the name Tribosphenida. Accordingly we propose that Tribosphenida be defined (consistent with traditional usage) as the first mammal possessing a dentition in which the protocone occludes with the talonid basin, plus all of its descendants: an apomorphy-based definition⁹. Holotheria¹⁰ refers to mammals characterized by a 'reversed triangles' molar pattern, that is, the ancestor of therians and *Kuehneotherium* plus all its descendants¹¹.

Synapsida

Holotheria (Wible *et al.* 1995, ref. 11)

Mammalia (Linnaeus 1758; *sensu* Rowe 1988, ref. 12)

Tribosphenida (see above)

Ambondro mahabo taxon nov. (Figs 2, 3)

Type specimen. UA-10602 (University of Antananarivo collection): mandibular fragment with three teeth, probably the ultimate pre-molar and the first two molars (Figs 2, 3).

Age, stratigraphic horizon and locality. See above. 16° 33' S, 46° 59' E, Ambondromamy region of the Mahajanga Basin, north-west Madagascar.

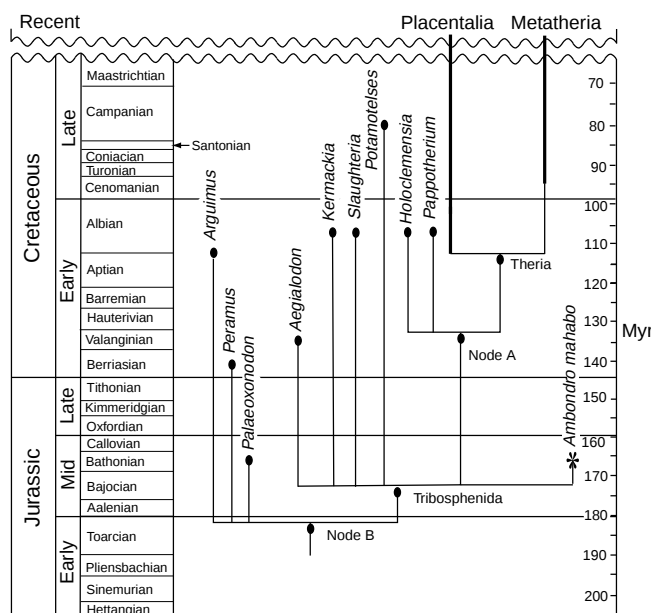


Figure 1 Chronology and phylogenetic relationships of selected tribosphenidans and near-holothere allies. Phylogeny distilled from a parsimony analysis²⁶ of 22 dental features¹⁵. Deltatheridians, *Ausktribosphenos* (both of contested phylogenetic placement)^{15,17,18,27}, *Vincelestes*²⁸ and various (generally poorly known) basal taxa have been omitted for clarity. The hypothetical common ancestor represented by Node B is characterized by expansion of the small distal blade or posterior spur of the lower molars (typical of early diverging holotheres) into an incipiently basined, cusp-bearing talonid⁸. Tribosphenida is characterized by multicusped, fully-basined talonids on the lower molars and an upper molar protocone^{8,15}—occurrence of the latter in *A. mahabo* is inferred from wear facets on the lower teeth (Figs 2, 3a, d). The hypothetical common ancestor represented by Node A is diagnosed (on the lower molars) by loss of the distal metacristid, labially placed attachment of cristid obliqua to posterior trigonid wall, close approximation of the lingual trigonid cusps to form an acute angle with the protoconid, and on upper molars by a preprotocrista extending labially past the paracone¹⁵.

Etymology. Referring to the village, Ambondromahabo, near which the type specimen was recovered.

Dental nomenclature. To simplify our diagnosis, a consideration of the tooth positions represented by UA-10602 is necessary. *Peramus tenuirostris*, the best known proximal outgroup of Tribosphenida, has eight postcanine teeth¹³, the identities of the fourth and fifth of which are debated^{8,14}. Although the anterior tooth in UA-10602 is possibly the first molar⁸, this tooth is most readily interpreted as the ultimate premolar (p-last), and the succeeding teeth as the first two molars (m1–2). Dental terminology is shown in Fig. 2.

Diagnosis. A diminutive holotherian mammal with strong lingual cingula on the p-last–m2, a basined talonid with weak distal metacristid on m1–2, and 'open' m1–2 trigonids (the three cusps describing an approximately 90° (that is, not acute) angle). A lack of various diagnostic attributes^{15,16} argues against metatherian affinities. The specimen differs from *Peramus* and all other Jurassic 'pre-tribosphenic' holotheres in its broader, more completely basined talonid, weaker distal metacristid, more labial intersection of cristid obliqua with posterior trigonid wall and strong lingual cingulum on the lower molars. *A. mahabo* is excluded from the clade stemming from Node A (Fig. 1) in retaining a distal metacristid and an 'open' trigonid; similarly, its lack of numerous autapomorphies rules out affinities to *Ausktribosphenos*¹⁷, an enigmatic (contested¹⁸ placental) holothere from the Early Cretaceous of Australia.

The most striking dental feature of *A. mahabo* is the basined talonid of m1–2 (slightly more elongated on m1). The talonids of these teeth, both very wide for a Jurassic mammal, are completely rimmed. A salient hypoconid is clearly seen on m1, as is its broken base on m2. A smaller hypoconulid occurs on m1; although it is well separated from the hypoconid, the two cusps are linked by a distinct cutting edge, a feature correlated with the presence of a functional metacone¹³. The little-worn entocristid of m1 is faintly swollen and papillate, but a distinct entoconid is lacking; two small cuspules occur on the entoconid of m2, but again an entoconid is absent. The cristid obliqua is continuous anteriorly with the labially shifted distal metacristid; the cristid obliqua thus meets the trigonid slightly more labially (on the labial side of the metaconid base) than in proximal outgroups to Tribosphenida (*Peramus*, *Palaeoxonodon*, *Arguimus*)—that is, below and slightly lingual to the notch in the metacristid, but still more lingually than is typical for therians and their nearest allies (node A, Fig. 1). The m1–2 talonid of *A. mahabo*

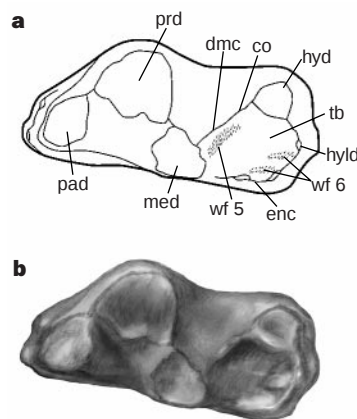


Figure 2 Posterior tooth (m2) of the holotype of *A. mahabo*. **a**, Dental terminology; **b**, occlusal view. Scale given in Fig. 3; anterior to the left. The m1–2 metaconids, damaged apically, show no evidence of the posterior accessory cusp present in various early tribosphenidans^{22,25} and proximal outgroups^{13–29}. The distal metacristid, weak but distinct, faces more labially than distally, and is continuous posteriorly with the cristid obliqua. The talonid basin is more extensively worn on m1. Abbreviations: co, cristid obliqua; dmc, distal metacristid; enc, entocristid; hyd, hypoconid; hyld, hypoconulid; med, metaconid; pad, paraconid; prd, protoconid; tb, talonid basin; wf, wear facet. Location of wear facets 5 and 6 are indicated.

is thus quite broad anteriorly relative to the incipient (lingually more open) basining in *Peramus*. Likewise, the talonid is broadened posteriorly over the condition in *Peramus* and other close tribosphenidan allies. The distal metacristid defines the lingual border of the hypoflexid, a groove for occlusion with the paracone. Unlike in later diverging tribosphenidans, this wear surface does not extend lingually across the back of the metaconid. Clear grooves occur lingual to the metacristid/cristid obliqua on the posterior trigonid face and anteriorly from the hypoconid (Fig. 3d), corresponding to wear facets five and six¹⁹—features generally associated with the presence of a protocone in the upper molars. Additional details of dental morphology are given in Figs 2 and 3.

Discussion. The presence of a distal trigonid crest (distal metacristid) in the therian outgroups *Potamotelses*, *Kermackia* and *Trinititherium*, the early tribosphenidan *Aegialodon* and some specimens of *Peramus* is functionally correlated with the absence of a paraconule and shelf-like preparaconule crista (=preprotocrista) mesial to the paracone. Retention of a weak distal metacristid in *A. mahabo* thus indicates primitive single-rank prevallum/postvallid shearing²⁰.

The Early Jurassic *Kuehneotherium* bears pronounced lingual cingulae on the lower molars, a condition not seen in basal holotheres outside Symmetrodonta and *Shuotherium*²¹. It is uncertain whether the condition in *A. mahabo* represents a retained primitive feature, or a secondary reversal (a strong lingual cingulum

arguably represents a primitive feature, given its occurrence in distant outgroups such as symmetrodonts, although more proximal outgroups, such as *Peramus* and *Palaeoxonodon*, lack the cingulum).

Discovery of *A. mahabo* confirms the proposal²² that “*Peramus* was a living fossil for its time, and that the tribosphenid innovation occurred in the upper Jurassic at the latest”. Although our find greatly extends the known range of the Tribosphenida, it is consistent with predicted ages for the earliest members of that clade based on the age of its closest known relatives (such as the Middle Jurassic *Palaeoxonodon*), and with molecular-clock analyses of amino-acid sequences indicating that monotremes and therians diverged before the Middle Jurassic (163–198 Myr ago)²³. The discovery of *A. mahabo* conflicts, however, with other molecular analyses advocating much more ancient divergences between major mammalian clades than supported by the fossil record³. For example, a Middle Jurassic (173 ± 12 Myr) divergence between Placentalia and Metatheria is proposed in ref. 3, but neither of these groups (nor even therians or members of the clade stemming from Node A of Fig. 1) is represented by fossils older than ~110 Myr (late Early Cretaceous). Although the fossil record of Mesozoic mammals is continually improving, dozens of Jurassic and Early Cretaceous holotherian mammals are known (some since 1838)²⁴, none of them of the kinds predicted by the molecular divergence estimates of ref. 3. Each new discovery like that reported here

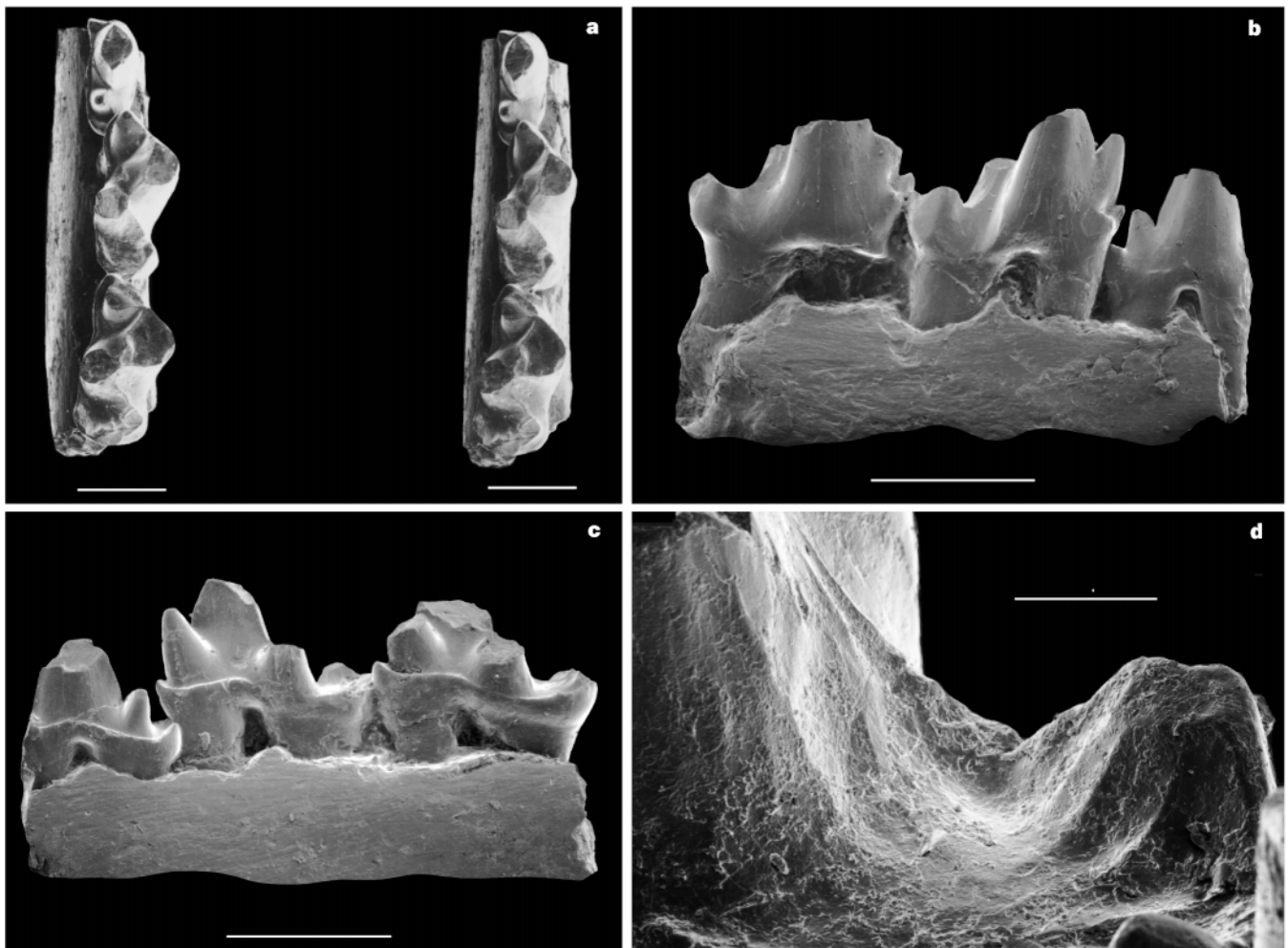


Figure 3 The holotype of *A. mahabo* in occlusal views. **a**, Stereopair; **b**, lingual; **c**, labial. **d**, Detail of m1 talonid in lingual view. Scale bars equal 1 mm in **a–c**, and 0.1 mm in **d**. The premolar is dominated by a large, anteromedian cusp that is flanked postero- (and presumably antero-) lingually by small accessory cusps. The posterior cusp is considerably smaller than the metaconids (cusp c) in typical (obtuse-angled)

symmetrodont molars; otherwise these teeth are similar. A strong cingulum is continuous across the lingual base of the crown, rimming the tooth posteriorly. The molar trigonids are distinctly less acutely angled (closed) than in therians and their near allies (such as *Holoclemensia*). The talonid is broad and distinctly basined; it is closed lingually by an entocristid, which continues anteriorly as a strong lingual cingulum.

reduces the already slim likelihood that the non-occurrence of therians before ~110 Myr is due to a preservational/sampling artefact.

Finally, in offering the first well-preserved holotherian dental remains known from the Middle to Late Jurassic of Gondwana, *A. mahabo* has important paleobiogeographic implications. Mesozoic mammal faunas of Laurasia and Gondwana were long considered to have remained isolated during much of the late Mesozoic; tribosphenidans (regarded as Laurasian in origin) were thought to have been excluded from the southern continents until latest Cretaceous or early Paleocene time, because Mesozoic tribosphenidans remained undiscovered in Gondwana⁴. To the contrary, recent finds in Morocco²⁵ and Australia¹⁷ indicate the occurrence of tribosphenidans on the southern continents as early as the earliest Cretaceous (that is, as early in Gondwana as in Laurasia). The discovery of *A. mahabo* inverts this biogeographic scenario even further, showing tribosphenidans to have appeared in Gondwana by the Middle Jurassic. They may have remained isolated there until earliest Cretaceous time, when they appear in North Africa, Europe and Asia. □

Received 15 March; accepted 28 June 1999.

1. Gradstein, F. M. et al. in *Geochronology, Time-Scales, and Global Stratigraphic Correlation* (eds Berggren, W. A., Kent, D. V. & Hardenbol, J.) 95–126 (Society of Sedimentary Geology Special Publication 54, 1995).
2. Krause, D. W., Prasad, G. V. R., von Koenigswald, W., Sahni, A. & Grine, F. E. Cosmopolitanism among Gondwanan Late Cretaceous mammals. *Nature* **390**, 504–507 (1997).
3. Kumar, S. & Hedges, S. B. A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920 (1998).
4. Bonaparte, J. F. Dispersal of Mesozoic mammals; new evidence. *Int. Geol. Cong. Abstr.* **28**, 170–171 (1989).
5. Flynn, J. J., Parrish, J. M., Wyss, A. R. & Simpson, W. F. New Triassic and Jurassic vertebrates from Madagascar. *J. Vert. Pal.* **18** (suppl.), 42A (1998).
6. Besairie, H. Géologie de Madagascar I. Les Terrains Sédimentaires. *Ann. Géol. Madagascar, Serv. des Mines* **35**, 1–463 (1972).
7. Bonaparte, J. F. in *The Beginning of the Age of Dinosaurs: Faunal Change Across the Triassic-Jurassic Boundary* (ed. Padian, K.) 247–258 (Cambridge Univ. Press, New York, 1986).
8. Clemens, W. A. & Lillegraven, J. A. in *Vertebrates, Phylogeny and Philosophy* (eds Flanagan, K. M. & Lillegraven, J. A.) 119–130 (*Contrib. Geol. Univ. Wyoming, Spec. Paper* 3, 1986).
9. de Queiroz, K. & Gauthier, J. Phylogeny as a central principle in taxonomy: Phylogenetic definitions of taxon names. *Syst. Zool.* **39**, 307–322 (1990).
10. Hopson, J. A. in *Major Features of Vertebrate Evolution, Short Courses in Paleontology* (eds Prothero, D. & Schoch, R. M.) 190–219 (Paleontological Soc., Knoxville, 1994).
11. Wible, J. R., Rougier, G. W., Novacek, M. J., McKenna, M. C. & Dashzeveg, D. A mammalian petrosal from the Early Cretaceous of Mongolia: implications for the evolution of the ear region and mammalian interrelationships. *Am. Mus. Novitates* **3149**, 1–19 (1995).
12. Rowe, T. Definition, diagnosis and origin of Mammalia. *J. Vert. Paleontol.* **8**, 241–264 (1988).
13. Clemens, W. A. & Mills, J. R. E. Review of *Peramus tenuirostris* Owen (Eupantotheria, Mammalia). *Brit. Mus. (Nat. Hist.) Bull. (Geol.)* **20**, 87–113 (1971).
14. McKenna, M. C. in *Phylogeny of the Primates: A Multidisciplinary Approach* (eds Luckett, W. P. & Szalay, F. S.) 21–46 (Plenum, New York, 1975).
15. Cifelli, R. L. in *Mammal Phylogeny: Mesozoic Differentiation, Multituberculates, Monotremes, Early Therians, and Marsupials* (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 205–215 (Springer, New York, 1993).
16. Cifelli, R. L. & de Muizon, C. Dentition and jaw of *Kokopellia juddi*, a primitive marsupial or near-marsupial from the medial Cretaceous of Utah. *J. Mamm. Evol.* **4**, 241–258 (1997).
17. Rich, T. H. et al. A tribosphenic mammal from the Mesozoic of Australia. *Science* **278**, 1438–1442 (1997).
18. Kielan-Jaworowska, Z., Cifelli, R. L. & Luo, Z. Alleged Cretaceous placental from down under. *Lethaia* **31**, 267–268 (1998).
19. Crompton, A. W. in *Early Mammals* (eds Kermack, D. M. & Kermack, K. A.) *Zool. Jour. Linn. Soc.* (suppl.) **50**, 65–87 (1971).
20. Fox, R. C. Molar structure and function in the Early Cretaceous Mammal *Pappotherium*: Evolutionary implications for Mesozoic Theria. *Can. J. Earth Sci.* **12**, 412–442 (1975).
21. Chow, M. & Rich, T. H. *Shuotherium dongi*, n. gen. and sp., a therian with pseudo-tribosphenic molars from the Jurassic of Sichuan, China. *Aust. Mammalogy* **5**, 127–142 (1982).
22. Sigogneau-Russell, D. & Ensom, P. Découverte, dans le Groupe de Purbeck (Berriasien, Angleterre), du plus ancien témoignage de l'existence de mammifères tribosphéniques. *C. R. Acad. Sci. Paris* **319**, 833–838 (1994).
23. Messer, M., Weiss, A. S., Shaw, D. C. & Westerman, M. Evolution of the Monotremes: Phylogenetic relationship to Marsupials and Eutherians, and estimation of divergence dates based on alpha-lactalbumin amino acid sequences. *J. Mamm. Evol.* **5**, 95–105 (1998).
24. de Blainville, H. Doutes sur le prétendu didelphe fossil de Stonesfield, ou à quelle classe, à quelle famille, à quel genre, doit-on rapporter l'animal auquel ont appartenu les ossements fossiles, à Stonesfield, désignés sous les noms de *Didelphis prevostii*, et *Didelphis bucklandii*, par les paléontologistes. *Comptes Rendus Acad. Sci. Paris* **VII**, 402–418 (1838).
25. Sigogneau-Russell, D. *Hypomylos phelizoni* nov. gen. nov. sp. une étape précoce de l'évolution de molaires tribosphéniques (Crétacé basal du Maroc). *Geobios* **25**, 461–483 (1992).
26. Swofford, D. L. PAUP 3.1.1 (Illinois Natural History Survey, Champaign, 1993).
27. Rougier, G. W., Wible, J. R. & Novacek, M. J. Implications of *Deltatheridium* specimens for early marsupial history. *Nature* **396**, 459–463 (1998).
28. Bonaparte, J. F. & Rougier, G. W. Mamíferos del Cretácico Inferior de Patagonia. *Memorias IV Congreso Latinoamericana de Paleontología (Santa Cruz de la Sierra, Bolivia)* **1**, 343–359.
29. Freeman, E. A Middle Jurassic mammal bed from Oxfordshire. *Palaeontology* **22**, 135–166 (1979).

Acknowledgements

This work was supported by the National Geographic Society, Field Museum of Natural History and the University of California. We thank the Ministre de l'Energie et des Mines of Madagascar for authorizing this work, S. Goodman and M. Ravokatra for help with our field efforts and G. Ravololonarivo Randria and A. Rasoamiaramanana. Field Museum volunteers D. Kinzig and R. Chisholm found this specimen during washing and microscope sorting of the matrix; K. Mazanek and M. Donnelly produced the illustrations; R. Andriantompohavana, L. Razafimanantso, A. Toto Volahy and A. Gandie aided us in the field; and W. Clemens and A. Smith provided advice.

Correspondence and requests for materials should be addressed to A.R.W. (e-mail: wyss@geology.ucsb.edu).

Transgenerational induction of defences in animals and plants

Anurag A. Agrawal^{*}, Christian Laforsch[†] & Ralph Tollrian[†]

^{*} Department of Entomology and Center for Population Biology, University of California at Davis, Davis, California 95616-8584, USA

[†] Ludwig-Maximilians-University München, Zoological Institute, Karlstrasse 25, 80333 München, Germany

Predators are potent agents of natural selection in biological communities. Experimental studies have shown that the introduction of predators can cause rapid evolution of defensive morphologies and behaviours in prey^{1–5} and chemical defences in plants^{6,7}. Such defences may be constitutively expressed (phenotypically fixed) or induced when predators initially attack^{8–10}. Here we show that non-lethal exposure of an animal to carnivores, and a plant to a herbivore, not only induces a defence, but causes the attacked organisms to produce offspring that are better defended than offspring from unthreatened parents. This transgenerational effect, referred to as a maternally induced defence, is in contrast to the more common defences induced in single individuals within a generation. Transgenerational induction of defences is a new level of phenotypic plasticity across generations that may be an important component of predator–prey interactions.

Inducible defences may provide an adaptive defensive strategy in which non-lethal cues from predators, herbivores or parasites provide a reliable indicator about the future risk of attack. In both animal and plant systems, inducible defences allow organisms to reap the benefits of defence while saving potential costs associated with investment in the defensive strategy when it is not needed. Such phenotypic benefits and costs have been demonstrated for inducible biochemical and morphological defences in plants, and for inducible morphological, life-history and behaviour defences in animals^{8–11}. If the parental environment predicts the quality of the progeny's environment, then parents may further enhance their net reproductive success by differentially endowing their offspring with phenotypes to cope with potential hazards such as predation. Such maternal effects have been reported in both animals and plants^{12–14}. These effects have traditionally been ascribed to passive consequences of the resource environment and have been viewed as phenotypic noise associated with thwarting adaptive evolution. More recently, however, maternal effects have been seen as being potentially adaptive^{12,15–18}. Adaptive maternal effects involve responding to information in the environment to improve the fitness of offspring by altering the offspring's phenotype. We have investigated the consequences of non-lethal exposure of wild radish plants (*Raphanus raphanistrum*) and waterfleas (*Daphnia cucullata*) to their predators, and have documented the effects on the defensive phenotypes of their offspring.

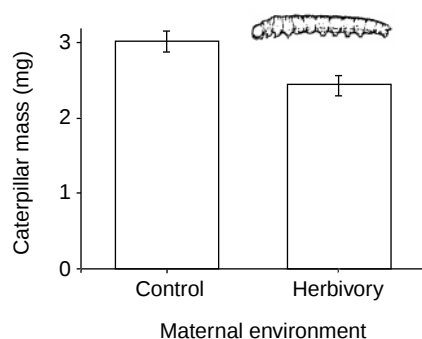


Figure 1 *Raphanus raphanistrum* resistance to herbivory as measured by growth of a specialist caterpillar, *Pieris rapae*, on the F_1 generation of seedlings from different maternal environments. Maternal plants were either subject to herbivory by *P. rapae* or left undamaged. This transgenerational induction of defences was not associated with differences in seed mass or seed concentrations of carbon or nitrogen. Error bars show s.e.

Wild radish plants damaged by a specialist caterpillar, *Pieris rapae*, induced tenfold higher concentrations of indole glucosinolates (mustard oil glycosides) and 30% higher densities of setose trichomes on newly formed leaves of damaged plants compared with undamaged control plants^{19,20}. The induction of these putative defences protected them against subsequent herbivory in several field experiments, and induced plants had 60% higher lifetime seed production than uninduced controls^{11,19}.

To examine the consequences of herbivory for the next generation of plants, we grew unmanipulated control plants and plants with 50% of each leaf consumed by a caged *P. rapae* larva, a natural herbivore of *R. raphanistrum*. Defensive glucosinolate profiles shifted in seeds from damaged plants compared with undamaged controls (multivariate analysis of variance; Wilks' $\lambda = 0.699$, $F_{4,26} = 2.803$, $P = 0.046$). Hydroxylated glucosinolates increased in concentration, whereas other classes of glucosinolates (aliphatic and indole) decreased. Herbivory on the maternal plant also induced changes in the number of trichomes per leaf in seedlings (A.A.A., unpublished observations). Seedling progeny from damaged and undamaged plants were challenged with *P. rapae* larvae. Caterpillars gained 20% less weight on seedlings whose parents were damaged than on seedlings whose parents were not damaged (Fig. 1, Table 1). This transgenerational induced plant resistance was not explained by seed mass variation or investment in primary metabolites (Table 1). Wild radish plants with induced defences produced seeds that did not differ in percentage composition of nitrogen or carbon compared with seeds from undamaged plants (nitrogen: control, 4.37 ± 0.07 (mean $\pm$ s.e.); maternal herbivory, 4.46 ± 0.06 ; $t = 0.988$, d.f. = 33, $P = 0.330$; carbon: control, 56.93 ± 0.30 ; maternal herbivory, 56.77 ± 0.19 ; $t = 0.454$, d.f. = 33, $P = 0.653$). Thus, induction of plant defences persisted in the progeny of infested plants. This effect may be the result of both a maternally induced defence and a greater rapidity of induction of plant defences in the offspring of damaged mothers.

Induced chemical defences have been best studied in plants, but induced morphological defences have been best studied in animals. Helmet formation in *Daphnia cucullata* (Fig. 2) is a textbook example of cyclomorphosis²¹ (seasonal variation in morphology),

Table 1 Effects of maternal environment (herbivory or no herbivory) and grandmaternal and maternal family on growth of caterpillars feeding on seedling plants

Source	d.f.	MS*	F	P
Maternal environment (ME)	1	1.503	7.516	0.009
Grandmaternal family (GF)	7	0.090	0.452	0.863
Maternal family nested within ME $\times$ GF	43	0.200	1.157	0.288
Seed mass	1	0.017	0.099	0.754
Error	74	0.173		

* MS, mean square.

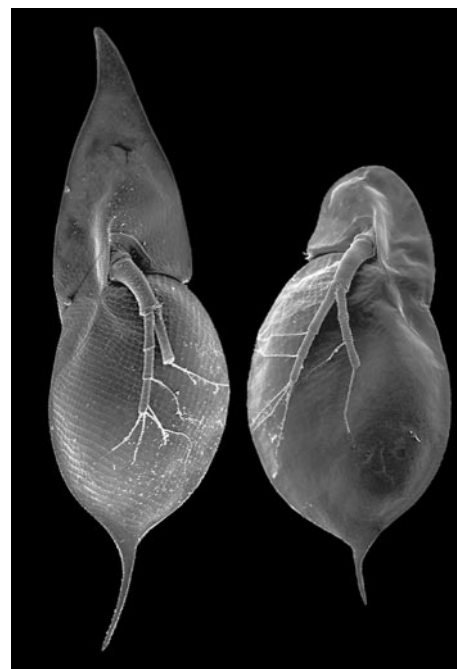


Figure 2 Scanning electron micrograph showing typical and predator-induced morphs of *Daphnia cucullata* of the same clone.

although the agents that maintain this polyphenism have not been adequately demonstrated^{22,23}. Morphological defences in cladocerans are induced by chemicals known as kairomones that are released by predators²⁴. In the presence of caged predaceous cladocerans, *Leptodora kindtii*, the relative helmet length of *D. cucullata* almost doubles (control, 15.53 ± 0.35 ; induced, 29.71 ± 0.49 ; $t = 23.72$, d.f. = 303, $P < 0.001$). Similar results were obtained for *D. cucullata* in the presence of caged aquatic larvae of the dipteran phantom midge, *Chaoborus flavicans* (control, 13.93 ± 0.15 ; induced, 27.88 ± 0.28 ; $t = 44.29$, d.f. = 470, $P < 0.001$). The induction of helmets acts as a defence by lowering the capture success by both predators: in controlled experiments, daphnids with induced helmets suffered lower mortality from both predators (mean ($\pm$ s.e.) number of prey captured per feeding trial; *Leptodora*: control, 6.43 ± 0.48 ; induced, 0.71 ± 0.42 ; $n = 7$, $P = 0.018$; *Chaoborus*: control, 1.86 ± 0.33 ; induced, 0.71 ± 0.22 ; $n = 14$, $P < 0.001$). These results indicate that a proximate cue used is a chemical released by predators, and defence against invertebrate predators is a major selective force favouring cyclomorphosis in *D. cucullata*.

To examine transgenerational induction of defences in *D. cucullata*, we raised daphnids in either control environments or environments with the *Chaoborus* kairomone. We then imposed four treatments and compared helmet size across the treatments in the F_1 and F_2 generations: (1) control daphnids (C), which, like their mothers, had not been in contact with the predator kairomone; (2)

Table 2 Helmet lengths of *Daphnia cucullata* progeny compared with ANOVA within each brood and age class

Source	Age	d.f.	MS	F	P
F_1 , first brood	Neonate	3,295	0.144, 0.004	40.51	<0.001
	Adult	3,239	0.603, 0.002	340.61	<0.001
F_1 , second brood	Neonate	3,266	0.170, 0.004	47.57	<0.001
	Adult	3,208	0.414, 0.002	193.44	<0.001
F_1 , third brood	Neonate	3,197	0.134, 0.002	71.65	<0.001
	Adult	3,176	0.297, 0.002	170.90	<0.001
F_2 , generation	Neonate	3,241	0.334, 0.003	132.11	<0.001
	Adult	3,215	1.202, 0.002	486.14	<0.001

MS and d.f. are shown first between groups and second within groups (error).

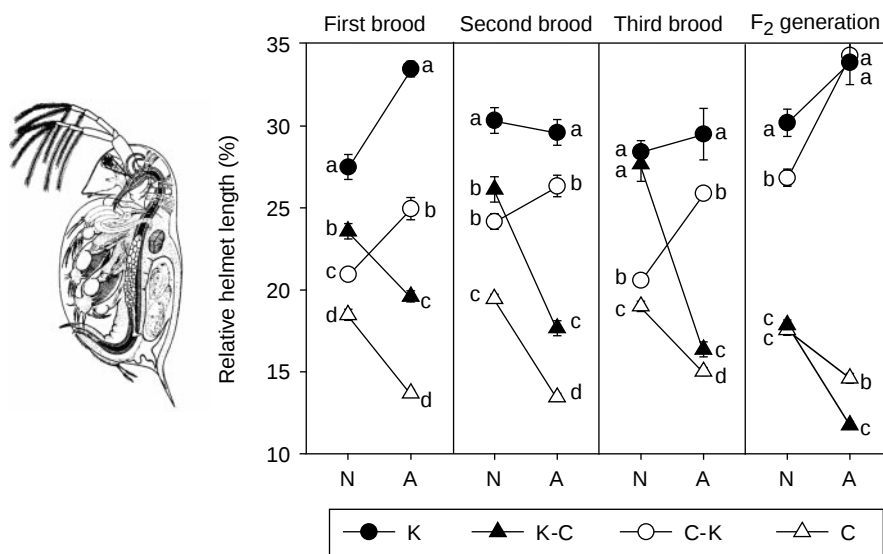


Figure 3 Relative helmet length (mean $\pm$ s.e.) of *Daphnia cucullata* (F₁ and F₂ generation) of four treatments, organized by brood number. K, kairomone treatment (*Chaoborus*); C, control; K $\rightarrow$ C, mother (F₀) had been transferred from kairomone to control treatment; C $\rightarrow$ K, mother (F₀) had been transferred from control to kairomone treatment. Mothers were transferred between environments after they had become pregnant. Daphnids are sensitive to the kairomone starting in a late embryo stage and in the juvenile stages (R.T., unpublished data). We analysed the first three broods of the F₀ mothers (for example, in

the transfer treatment, three broods following the transfer). Thus, only the first brood shared the maternal environment for a very short time. Lines between neonates (N) and adults (A) show relative helmet growth within each treatment. Results of each overall ANOVA are shown in Table 2. Homogeneous groups within each brood and age class are indicated by vertical ordered letters (Bonferroni adjusted). The F₂ generation is the first brood offspring of the F₁ first brood. Where error bars are not shown, they were smaller than the symbol.

kairomone daphnids (K), which, like their mothers, had been in permanent contact with the kairomones; (3) daphnids which had been born from control mothers and were transferred to the kairomone environment (C $\rightarrow$ K); and (4) daphnids which had been born from kairomone mothers and were transferred to the control environment (K $\rightarrow$ C). The latter two treatments were used to distinguish between a maternally induced defence and environmental induction in the progeny. Pregnant mothers were transferred between environments shortly after they had deposited their first clutch into their brood pouch. At this stage, young daphnids are not yet sensitive to the kairomone (R.T., unpublished data).

F₁ neonates varied in helmet length between the four treatments (Table 2, Fig. 3), with offspring from mothers in the kairomone environment always having larger helmets than offspring from mothers in the control environment. Significant differences between daphnids with the same F₁ environment but different maternal environments (compare C with K $\rightarrow$ C, and K with C $\rightarrow$ K) demonstrate the importance of the maternal environment in helmet formation. This pattern persisted in subsequent broods that were initiated after the mothers had been transferred, which demonstrates the lasting influence of the maternal environment on the transgenerational induction of helmet defences (Fig. 3). A comparison of treatments with the same maternal environment but different offspring environments (compare K with K $\rightarrow$ C, and C with C $\rightarrow$ K) indicates that the late embryonic environment also had an influence on helmet induction, albeit a smaller effect than that of maternal environment (Fig. 3).

Postnatal helmet development in *D. cucullata* is induced by the juvenile environment, as can be seen from the interactions between helmet length and age (Fig. 3). The K $\rightarrow$ C daphnids, which were born with large helmets, reached maturity with relatively small helmets. In contrast, neonates from C $\rightarrow$ K mothers were born with small helmets but reached maturity with relatively large helmets. However, the juvenile environment could not induce phenotypes that matched those of maternally induced individuals. For example, the helmet size of adult daphnids from the kairomone treatment (K) could not be reached by C $\rightarrow$ K daphnids, which also spent their whole lives in contact with the kairomone. Similarly, K $\rightarrow$ C

daphnids reached maturity with larger helmets than control (C) daphnids.

The F₂ generation from C $\rightarrow$ K daphnids had significantly smaller helmets than did K daphnids, indicating that grandmaternal environment constrained the size of helmets in the F₂ neonates (Fig. 3). However, the F₂ generation from K $\rightarrow$ C daphnids did not have a different helmet size than C daphnids. In both cases, the maternal (F₁) environment influenced helmet induction in the F₂ generation.

The greater defences of organisms whose parents were threatened by predation provides a case of adaptive phenotypic plasticity across generations. Adaptive maternal effects are distinct from other adaptations, however, in that the genes responsible are activated in the maternal generation, whereas the phenotypes on which natural selection may act are expressed in the progeny. This lagged link between parental environment and offspring phenotype may amplify or subdue responses to natural selection^{18,25}.

Transgenerational effects may provide tools to overcome developmental constraints acting on individuals at critical stages in the life cycle. For example, the later development of a maximally large helmet for initially short-headed daphnids might not be possible, but transgenerational effects imposed by the maternal environment may allow for such maximal defences. Furthermore, progeny born with a maternally induced defence avoid a major disadvantage of inducible defences: that the defence is not initially present when needed (there is a lag phase²⁶). The immediate development of plant resistance in wild radish seedlings is important because the establishment phase has a strong influence on plant fitness^{11,27}. Maternally induced progeny will be just as well protected as permanently defended individuals, and better protected than undefended individuals or individuals that respond to threats early in life. Because predation can be such a strong selective force, we posit that transgenerational induced defences are a logical extension of fitness-enhancing strategies of organisms under threat. Several other systems are strong candidates for having maternally induced defences^{28–30}. That such effects are present in both animals and plants reinforces the view that the expression of adaptive phenotypes in prey may be driven by predation from higher trophic levels. □

Methods

Wild radish experiments.

We used *R. raphanistrum* seeds from a second generation of untreated greenhouse-grown plants. We germinated about 10 seeds from each of 13 maternal families in a greenhouse. At the four-leaf stage, each plant was randomly assigned to one of the two treatments (3–5 plants per treatment per family). The caterpillar herbivory treatment was maintained throughout the growth of the plant. Effects of herbivory on seed set are reported elsewhere¹⁹. Seeds from 8 of the original 13 families were chosen for the transgenerational experiment because they spanned the range of tolerance to herbivory¹⁹, and 2–4 seeds from each of 58 maternal plants within 8 grandmaternal families (a total of 126 plants) were grown in a greenhouse and inoculated with a single newly hatched *P. rapae* larva. Caterpillars were not caged. After four days, the caterpillars were weighed.

Effects of grandmaternal family (fixed), maternal environment (control or herbivory; fixed), maternal family (nested within grandmaternal family by treatment interaction; random) and seed mass (covariate) on caterpillar growth were analysed using a mixed-model analysis of variance (ANOVA). *F*-ratios for grandmaternal family and treatment effects were calculated with maternal family nested in grandmaternal family by treatment interaction mean-square and degrees of freedom in the denominator.

For phytochemical analysis, glucosinolates were analysed using modified procedures for determination of trimethylsilyl glucosinolate derivatives with capillary gas chromatography and flame ionization detection²⁰. Seed nitrogen and carbon were determined from a separate set of seeds using dynamic flash-combustion and gas-chromatographic separation and a thermal-conductivity detection system (Division of Agriculture and Nature Resources, University of California, Davis).

Daphnia experiments.

We used a *Daphnia cucullata* clone isolated from Thaler lake, Germany. The experiments were conducted in the laboratory under constant conditions at 20 °C and fluorescent light in a synthetic medium. The *F*₀ generation had been synchronized by always raising the third brood offspring born within 12 h, starting from a single *Daphnia*. *Daphnia* were fed *ad libitum* daily with *Scenedesmus acutus* (1.5 mg C l⁻¹).

Animals with freshly deposited eggs were placed into 0.75 l medium containing a 125-µm net cage, enclosing either 10 fourth-instar larvae of *Chaoborus flavicans* or 4 *Leptodora kindtii*, or no predators for controls. The net cages prevented direct contact between predators and prey.

Predation experiments with *Chaoborus* were conducted for 0.5 h with each prey morph separately in 100 ml medium with 10 prey organisms and a single predator. Predation trials with *Leptodora* were conducted with 10 prey organisms of each morph together in 0.5 l in 24-h experiments. In each experiment, animals of the same age and body size class (0.6–0.8 mm, from the eye to the base of the tail spine) but of different morphology were compared.

We induced helmet formation in *Daphnia* by using *Chaoborus* kairomones. We placed four *Chaoborus* larvae in net cages in 1.5-l beakers and changed the water every day. *Chaoborus* were fed daily with 10–15 prey (*D. cucullata* and *Ceriodaphnia* sp.). The *F*₀ generation had been born and raised in these beakers in either the control or kairomone treatment. We used three beakers per treatment as replicates, which did not differ and were pooled for analysis.

Daphnids were measured with a digital image-analysis system (SIS, Münster, Germany). To compensate for small changes in body length within an age class, we calculated the relative helmet length (helmet length/body length × 100). Relative helmet length is a good predictor of the defensive effect within an age class. The relative values were arcsin-transformed for analysis. In the induction treatment, differences between control and kairomone treatment were compared using *t*-tests. Effects of induced *Daphnia* phenotypes on predation were analysed by using a Mann–Whitney–*U*-test for *Chaoborus* and with a paired Wilcoxon test for related samples for *Leptodora*. To test for transgenerational effects, we compared the treatments within each brood and age class (Table 2, Fig. 3) with ANOVA and Bonferroni adjustments.

Received 29 March; accepted 28 June 1999.

- Thompson, J. N. Rapid evolution as an ecological process. *Trends Ecol. Evol.* **13**, 329–332 (1998).
- Seeley, R. H. Intense natural selection caused a rapid morphological transition in a living marine snail [*Littorina obtusata*]. *Proc. Natl Acad. Sci. USA* **83**, 6897–6901 (1986).
- Magurran, A. E., Seghers, B. H., Carvalho, G. R. & Shaw, P. W. Behavioural consequences of an artificial introduction of guppies (*Poecilia reticulata*) in N. Trinidad: Evidence for the evolution of anti-predator behaviour in the wild. *Proc. R. Soc. Lond. B* **248**, 117–122 (1992).
- Reznick, D. N., Shaw, F. H., Rodd, F. H. & Shaw, R. G. Evaluation of the rate of evolution in natural populations of guppies (*Poecilia reticulata*). *Science* **275**, 1934–1937 (1997).
- Lenski, R. E. & Levin, B. R. Constraints on the coevolution of bacteria and virulent phage: A model, some experiments, and predictions for natural communities. *Am. Nat.* **125**, 585–602 (1985).
- Thompson, J. N. & Burdon, J. J. Gene-for-gene coevolution between plants and parasites. *Nature* **360**, 121–125 (1992).
- Frita, R. S. & Simms, E. L. (eds) *Plant Resistance to Herbivores and Pathogens* (Univ. Chicago Press, 1992).
- Tollrian, R. & Harvell, C. D. (eds) *The Ecology and Evolution of Inducible Defenses* (Princeton Univ. Press, Princeton, NJ, 1999).
- Agrawal, A. A., Tuzun, S. & Bent, E. (eds) *Induced Plant Defense against Pathogens and Herbivores: Biochemistry, Ecology, and Agriculture* (American Phytopathological Society Press, St Paul, Minnesota, in the press).
- Karban, R. & Baldwin, I. T. *Induced Responses to Herbivory* (Univ. Chicago Press, 1997).
- Agrawal, A. A. Induced responses to herbivory and increased plant performance. *Science* **279**, 1201–1202 (1998).
- Mousseau, T. A. & Fox, C. W. (eds) *Maternal Effects as Adaptations* (Oxford Univ. Press, New York, 1998).
- Rossiter, M. Incidence and consequences of inherited environmental effects. *Annu. Rev. Ecol. Syst.* **27**, 451–476 (1996).

- Roach, D. A. & Wulff, R. D. Maternal effects in plants. *Annu. Rev. Ecol. Syst.* **18**, 209–236 (1987).
- Schmitt, J., Niles, J. & Wulff, R. D. Norms of reaction of seed traits to maternal environments in *Plantago lanceolata*. *Am. Nat.* **139**, 451–466 (1992).
- Tollrian, R. Predator-induced morphological defenses: Costs, life history shifts, and maternal effects in *Daphnia pulex*. *Ecology* **76**, 1691–1705 (1995).
- Fox, C. W., Thakar, M. S. & Mousseau, T. A. Egg size plasticity in a seed beetle: An adaptive maternal effect. *Am. Nat.* **149**, 149–163 (1997).
- Thiede, D. A. Maternal inheritance and its effect on adaptive evolution: A quantitative genetic analysis of maternal effects in a natural plant population. *Evolution* **52**, 998–1015 (1998).
- Agrawal, A. A. Induced responses to herbivory in wild radish: Effects on several herbivores and plant fitness. *Ecology* **80**, 1713–1723 (1999).
- Agrawal, A. A., Strauss, S. Y. & Stout, M. J. Costs of induced responses and tolerance to herbivory in male and female fitness components of wild radish. *Evolution* (in the press).
- Hutchinson, G. E. *A Treatise on Limnology* Vol. 2 (Wiley, New York, 1967).
- Jacobs, J. in *Daphnia* (eds Peters, R. H. & de Bernardi, R.) 325–252 (Memorie dell'Istituto Italiano di Idrobiologia Dott. Verbania Palanza, 1987).
- Tollrian, R. Predator-induced helmet formation in *Daphnia cucullata* (Sars). *Arch. Hydrobiol.* **119**, 191–196 (1990).
- Tollrian, R. & Von Elert, E. Enrichment and purification of *Chaoborus* kairomone from water: Further steps toward its chemical characterization. *Limnol. Oceanogr.* **39**, 788–796 (1994).
- Schluter, D. & Gustafsson, L. Maternal inheritance of condition and clutch size in the collared flycatcher. *Evolution* **47**, 658–667 (1993).
- Lenski, R. E. et al. Epistatic effects of promoter and repressor functions of the Tn10 tetracycline-resistance operon on the fitness of *Escherichia coli*. *Mol. Ecol.* **3**, 127–135 (1994).
- Stanton, M. L. Seed variation in wild radish: Effect of seed on components of seedling and adult fitness. *Ecology* **65**, 1105–1112 (1984).
- Gilbert, J. J. Rotifer ecology and embryological induction. *Science* **151**, 1234–1237 (1966).
- Roberts, D. A. Acquired resistance to Tobacco Mosaic Virus transmitted to the progeny of hypersensitive tobacco. *Virology* **124**, 161–163 (1983).
- Shine, R. & Downes, S. J. Can pregnant lizards adjust their offspring phenotypes to environmental conditions? *Oecologia* **119**, 1–8 (1999).

Acknowledgements

We thank R. Karban, the plant-herbivore group at Davis and W. Gabriel for advice and encouragement; M. Morra and V. Borek for help with phytochemical analyses; R. J. Mercader, M. Kredler, S. Y. Strauss, J. Kniskern and E. Hochmuth for help with experiments; and L. S. Adler, S. Diehl, H. Dingle, R. Karban, R. E. Lenski, T. W. Schoener, S. Y. Strauss, J. S. Thaler, D. A. Thiede, and T. G. Whitham for comments on the manuscript. This work was supported by the US NSF.

Correspondence and requests for materials should be addressed to A.A.A. (e-mail: aaagrawal@ucdavis.edu) or R.T. (e-mail: Ralph.Tollrian@lrz.uni-muenchen.de).

Genetic enhancement of learning and memory in mice

Ya-Ping Tang*, Eiji Shimizu*, Gilles R. Dube†, Claire Rampon*, Geoffrey A. Kerchner‡, Min Zhuo‡, Guosong Liu† & Joe Z. Tsien*

* Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544-1014, USA

† Department of Brain and Cognitive Science, The Center for Learning and Memory, MIT, Cambridge, Massachusetts 02139, USA

‡ Departments of Anesthesiology and Neurobiology, School of Medicine, Washington University, St Louis, Missouri 63110, USA

Hebb's rule (1949) states that learning and memory are based on modifications of synaptic strength among neurons that are simultaneously active. This implies that enhanced synaptic coincidence detection would lead to better learning and memory. If the NMDA (N-methyl-D-aspartate) receptor, a synaptic coincidence detector^{1–4}, acts as a graded switch for memory formation, enhanced signal detection by NMDA receptors should enhance learning and memory. Here we show that overexpression of NMDA receptor 2B (NR2B) in the forebrains of transgenic mice leads to enhanced activation of NMDA receptors, facilitating synaptic potentiation in response to stimulation at 10–100 Hz. These mice exhibit superior ability in learning and memory in various behavioural tasks, showing that NR2B is critical in gating the age-dependent threshold for plasticity and memory formation. NMDA-receptor-dependent modifications of synaptic

efficacy, therefore, represent a unifying mechanism for associative learning and memory. Our results suggest that genetic enhancement of mental and cognitive attributes such as intelligence and memory in mammals is feasible.

The NMDA receptors are heteromeric complexes consisting of subunit 1 (NR1) and various NR2 subunits^{5,6}. The NR1 subunit is essential for channel function, whereas the NR2 subunit regulates channel gating and Mg^{2+} dependency⁷. In adult forebrain regions such as the hippocampus and cortex, only NR2A and NR2B subunits are available to form the receptor complex with NR1 subunit. The recombinant NR1–NR2B complex *in vitro* shows longer excitatory postsynaptic potentials (EPSPs) than does the NR1–NR2A complex⁸. Increased incorporation of NR2B into the receptor complex *in vivo* should allow the NMDA receptors to increase the time window for detecting synaptic coincidence. NR2B expression is downregulated during the transition from juvenile to adult^{9,10}, correlating with the gradual shortening of the EPSP duration of the NMDA channel^{11,12}. This could decrease NMDA-mediated plasticity, and perhaps explain decreased memory performance in adult animals including songbirds, monkeys and humans^{13,14}.

To test whether the NR2B subunit is crucial for implementing Hebb's rule and gating synaptic plasticity and memory, we over-expressed the NR2B subunit postnatally in the mouse forebrain using the CaM-kinase-II promoter^{15,16} (Fig. 1a). Of seven lines produced, we report here results from two independent lines (Tg-1 and Tg-2) that we have systematically analysed. They show similar expression patterns and levels of NR2B, and nearly identical electrophysiological and behavioural phenotypes.

These transgenic animals, named *Doogie*, show normal growth and body weights and mate normally. Their open-field behaviour is indistinguishable from that of wild-type littermates, and we observed no seizure or convulsion in transgenic animals. As our northern blot analysis indicates, NR2B transgene expression is enriched in the cortex and hippocampus, with little expression in the thalamus, brainstem and cerebellum (Fig. 1b). Western blot analysis showed about twice as much NR2B protein in the cortex and hippocampus of transgenic mice as in the wild type (Fig. 1c). There is also a slight increase in NR1 protein, but no change in NR2A expression in these regions (Fig. 1c). This indicates that both the ratio of NR2B to NR2A in the receptor complex and the total number of NMDA receptors may be increased. We investigated the transgene's anatomical distribution using *in situ* hybridization and found that the transgene was highly enriched in the cortex, striatum, hippocampus and amygdala (Fig. 1d). Light microscopy showed no gross structural abnormalities in the transgenic animals (Fig. 1e, f). The shapes and architecture of dendritic spines in the hippocampus and cortex are also normal (Fig. 1g).

To evaluate the elementary properties of the NMDA receptors in single synapses, we used a single-bouton recording technique¹⁷. Using FM 1-43 dye to label functional synaptic sites in cultured hippocampal neurons, we positioned the tip of an iontophoretic electrode containing 150 mM glutamate next to a relatively isolated synapse (Fig. 2a) and applied glutamate to determine the functional properties of glutamate receptors located in that synapse. Glutamate-evoked responses consisted of either AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) current alone or both AMPA and NMDA currents, depending on the cell potential. The NMDA component was identified by its 'J'-shaped current–voltage relationship and long decay time (Fig. 2b). NMDA currents were isolated by clamping the cells to +40 mV to remove the voltage-dependent Mg^{2+} block. Initial experiments were carried out with 5 μ M DNQX in the bath solution to block AMPA receptors, but we subsequently omitted the antagonist because NMDA current could clearly be isolated from AMPA current based on their respective time courses. All the experiments were conducted in blind fashion.

We first determined the glutamate dose–response curve of the

synaptic NMDA receptors (an index of the glutamate affinity of the NMDA receptor) and its voltage-dependent Mg^{2+} block, and found no difference between the two groups of mice (data not shown). As the recombinant NR2B subunit determines the decay phase of NMDA currents *in vitro*⁸, we next measured the NMDA-channel decay time for currents evoked by a saturating dose of glutamate (100 nA of iontophoretic current, as determined from the dose–response relationship) in both transgenic and wild-type neurons. Although we found no difference in decay time at day 10 or 14 *in vitro*, the decay time of the NMDA currents from transgenic neurons at day 18 was 1.8-fold longer than that of wild-type neurons (Fig. 2c, inset and Fig. 2e; $P < 0.005$). In addition, transgenic neurons retained the juvenile-like, single-synapse peak NMDA-current amplitude over time in culture. This is in contrast to that of wild-type neurons, which decreased significantly by day 18 *in vitro* (Fig. 2d, significance between NR2B ($n = 18$) and wild type ($n = 8$), $P < 0.01$; significance between day 14 ($n = 8$) and day 18 ($n = 8$) for wild type, $P < 0.01$). This indicates that the total number of NR2B-containing NMDA receptors per single synapse is also increased in transgenic neurons at this stage. These age-dependent changes in channel decay time and peak amplitude are consistent with the *in vivo* observation that the NR2B transgene messenger RNA was detectable but low at postnatal day 14 (P14), and gradually increased to steady level about one week later (data

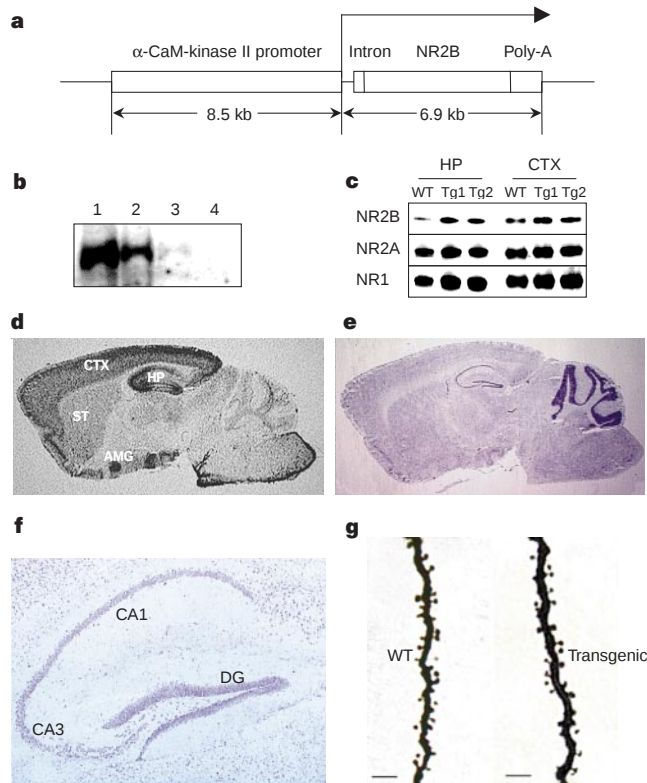


Figure 1 Construction and biochemical characterization of transgenic NR2B mice. **a**, The construct pJT–NR2B for production of NR2B transgenic mice. kb, kilobases.

b, Expression of NR2B transgene mRNA in transgenic mice. Lane 1, cortex/striatum/amygdala; lane 2, hippocampus; lane 3, brain stem and thalamus; lane 4, cerebellum. **c**, Synaptic NMDA-receptor protein in hippocampus (HP) and cortex (CTX) in both transgenic lines (Tg-1 and Tg-2) and wild-type (WT). The same immunoblot was used for blotting with antibodies against NR1 (relative molecular mass 120K), NR2A (170K) and NR2B (180K). **d**, Forebrain-specific expression of NR2B transgene revealed by *in situ* hybridization. CTX, cortex; ST, striatum; HP, hippocampus; AMG, amygdala. **e**, Normal brain morphology in transgenic mice (Nissl staining). **f**, Higher magnification of the Nissl-stained transgenic hippocampus. DG, dentate gyrus; CA1 and CA3 are marked. **g**, Golgi staining of the dendritic spines of CA1 cells from wild-type (left) and transgenic mice (right). Scale bar, 5 μ m.

not shown).

The prolonged decay time and the maintenance of a large single-synapse peak amplitude should result in increased charge transfer through the synaptic NMDA-receptor channel. We calculated the total amount of charge transfer associated with the activation of the NMDA receptors at a single synapse by integrating the area under the recorded current curve between the time of glutamate application and 400 ms later. The results clearly indicate that, at day 18 *in vitro*, the total charge transfer through single-synapse NMDA receptors was about 4-fold larger in transgenic mice than in controls (2.5 ± 0.7 pC ($n = 8$) in wild-type versus 9.8 ± 1.7 pC ($n = 18$) in NR2B mice, $P < 0.001$; Fig. 2f). Therefore, overexpression of the NR2B transgene results in prolonged opening of the NMDA receptors for detecting coincidence and enhanced NMDA activation in individual synapses, thus retaining several features of juvenile NMDA-receptor properties.

As blockade of the NMDA receptors prevents the induction of major forms of plasticity such as long-term potentiation and long-term depression (LTP and LTD)^{2,3}, we examined whether the

increased window for coincidence detection by NMDA receptors leads to enhanced synaptic plasticity in the CA1 region of the hippocampus. Using hippocampal slices prepared from 4–6-month-old animals, we first measured (in blind fashion) NMDA-mediated field EPSPs in both adult wild-type and transgenic mice in the Schaffer collateral CA1 path. NMDA receptor-mediated EPSPs were isolated in the presence of $10 \mu\text{M}$ CNQX and 0.1 mM Mg^{2+} . NMDA-receptor-mediated field EPSPs in transgenic mice were significantly greater than those in wild-type mice, indicating that the overexpression of NR2B enhanced NMDA-receptor-mediated field responses (Fig. 3b). In addition, we confirmed that the observed synaptic responses were NMDA-receptor-dependent by showing that they were sensitive to the NMDA-receptor antagonist D(-)-amino-7-phosphonovaleric acid (AP5; $100 \mu\text{M}$; $n = 3$; data not shown).

We then studied AMPA-receptor-mediated responses and found no difference in AMPA-mediated field EPSPs between transgenic mice ($n = 62$ slices/14 mice) and their wild-type littermates ($n = 50$ slices/16 mice, data not shown). Furthermore, paired-

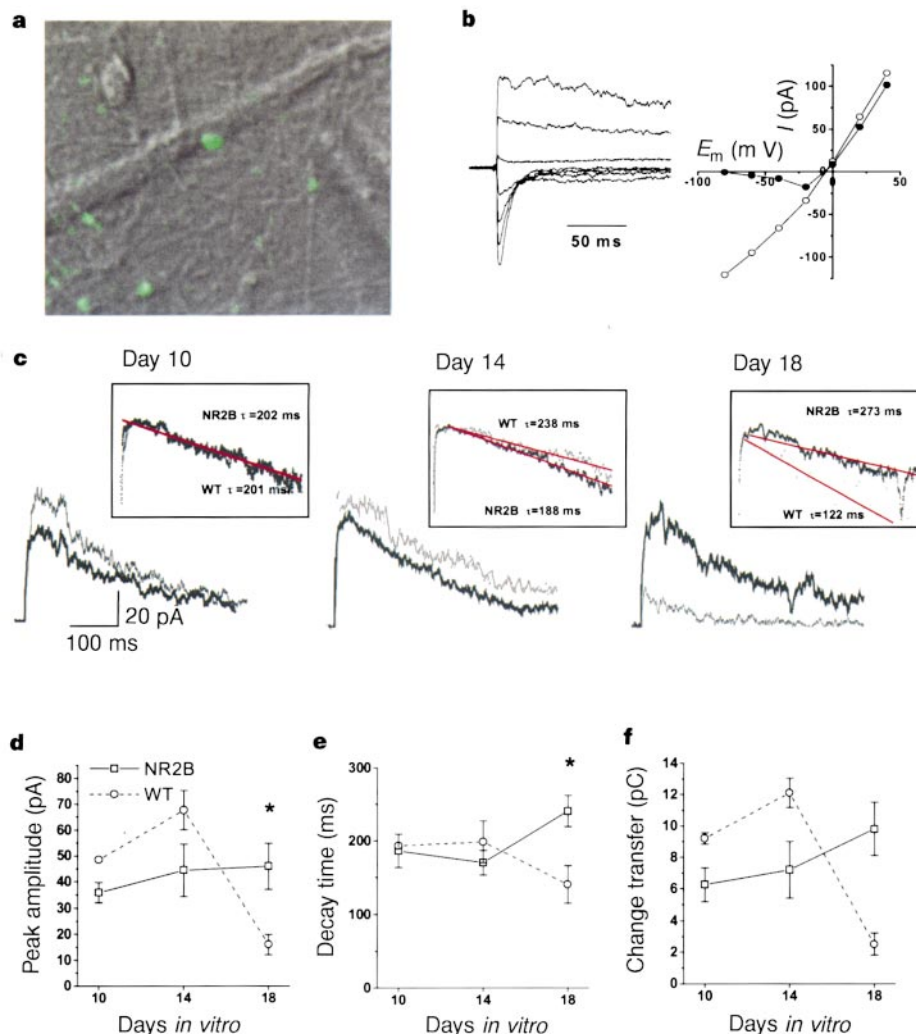


Figure 2 Developmental changes in NMDA current at single synapse. **a**, Confocal image of a dendrite with single synapses marked by FM 1-43 (green). The iontophoretic electrode on the right (out of the focal plan) was brought into close proximity to the FM spot to deliver glutamate (1 ms). **b**, Representative example of a current-voltage relationship of glutamate-evoked response from a single synapse from a wild-type neuron. At -80 to -40 mV, only non-NMDA current was observed, whereas at more positive potentials both non-NMDA (peak current typically observed at 3 ms after application) and NMDA current (peak current measured 30–40 ms after application) were recorded. The proportion of NMDA current evoked displays a typical 'J' shaped I - V relation (black circles) whereas non-NMDA current varies linearly with the membrane potential (open circles).

c, Representative examples of NMDA currents at $+40$ mV recorded from WT (light trace) and transgenic mice (dark trace) at days 10, 14 and 18 *in vitro*, respectively. Insets display the same traces, normalized and expressed as a semi-log plot to emphasize the decay portion of NMDA currents. A single exponential, which provided excellent fits, was used to assess the decay time τ and values for the representative traces are indicated. **d–f**, Averaged values for peak amplitude, decay time and charge transfer of NMDA currents, respectively. Each point represents mean $\pm$ s.e.m. of 8–18 experiments obtained from 18 synapses on 13 neurons from 6 wild-type mice and 31 synapses on 19 neurons derived from 9 transgenic (Tg-1 and Tg-2 mice). Asterisks, significance between WT and transgenic mice ($P < 0.01$, 2-tailed unpaired Student's *t*-test).

pulse facilitation (PPF), which gives an indication of presynaptic function, was similar in transgenic ($n = 11$ slices/7 mice) and wild-type hippocampal slices ($n = 7$ slices/5 mice) (Fig. 3a). These results indicate that both presynaptic function and postsynaptic AMPA receptors are normal in transgenic animals.

To assess bidirectional synaptic plasticity in the 1–100 Hz range¹⁸, we conducted a series of LTP/LTD experiments in the Schaffer collateral CA1 pathway in blind fashion. A single tetanic stimulation (100 Hz, 1 s) typically evoked smaller but reliable potentiation in 4–6-month-old control slices in comparison to that evoked in younger adult slices. However, the same stimulus evoked significantly larger potentiation in transgenic slices (Fig. 3c; transgenic, $n = 9$ slices/6 mice; wild-type, $n = 10$ slices/8 mice). The enhancement of potentiation was not due to changes in inhibitory GABA (γ -aminobutyric acid)-mediated mechanisms as it was also observed in the presence of $100 \mu\text{M}$ picrotoxin (transgenic, $n = 4$ slices/3 mice, mean $241.1 \pm 48.2\%$; wild-type, $n = 5$ slices/5 mice, mean $140.3 \pm 19.1\%$; $P < 0.05$ compared to transgenic mice). Moreover, the enhanced LTP was completely blocked by AP5

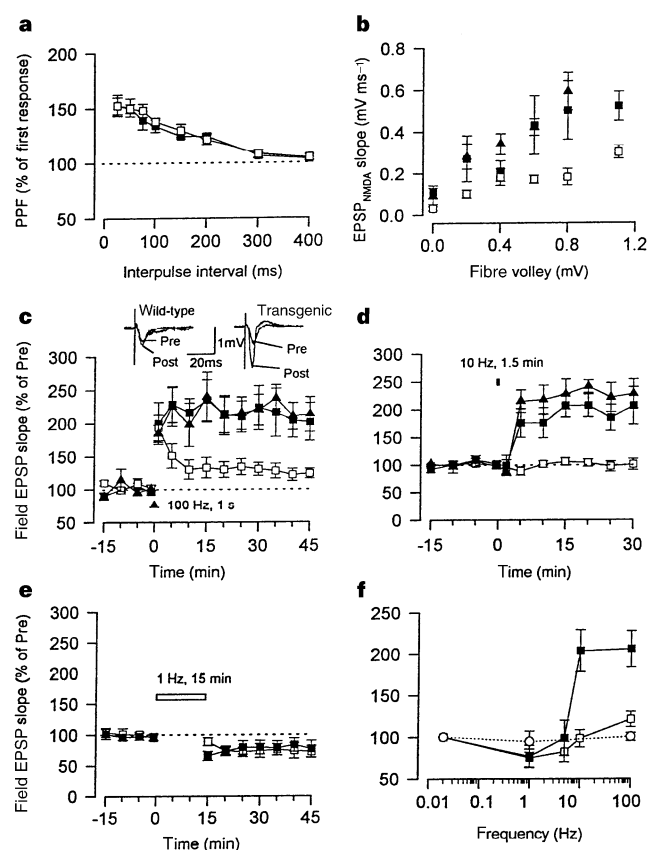


Figure 3 Selective enhancement of 10–100-Hz-induced potentiation in transgenic mice. Tg-1, filled squares; Tg-2, filled triangles; wild-type, open squares. **a**, Wild-type and transgenic mouse slices showed no significant difference in PPF of the EPSP at various interpulse intervals. **b**, Transgenic slices had larger NMDA-receptor-mediated EPSPs than wild-type slices. At 1.1 mV fibre volley, the area under the EPSP_{NMDA} was 124.8 ± 20.6 ($\text{mV} \times \text{ms}$) in transgenic mice and 31.1 ± 5.3 ($\text{mV} \times \text{ms}$) in controls ($P < 0.001$). **c**, Tetanic stimulation (100 Hz, 1 s) induced significantly more potentiation in transgenic slices (Tg-1, 9 slices/6 mice; Tg-2, 6 slices/3 mice) than in wild-type slices ($n = 10$ slices/8 mice); Inset: representative records of the EPSP before (Pre) and 45 min after (Post) tetanus in a WT (left) and transgenic (right) slice. **d**, 10-Hz stimulation for 1.5 min produced significant synaptic potentiation in transgenic slices (Tg-1, 5 slices/5 mice; Tg-2, 4 slices/3 mice) but not WT slices (9 slices/9 mice). **e**, LTD induced by low-frequency stimulation (1 Hz for 15 min) is similar between WT and transgenic slices. **f**, Summary data for synaptic plasticity at different frequencies. For comparisons, results from our previous study of CA1-specific NR1 knockout mice (open circles, dotted line) is included.

($100 \mu\text{M}$; data not shown). In addition, in contrast to the normal fast decay of the field EPSP in control slices during the first 10 min after tetanic stimulation, there was no decay at all. Instead, it remained maximally potentiated, a feature resembling the juvenile LTP observed in postnatal day 15 animals¹⁹.

Enhanced long-term potentiation in the transgenic slices was also observed when we applied prolonged, repetitive stimulation (10 Hz) to the Schaffer–CA1 path. Although 10-Hz stimulation for 1.5 min (900 pulses) did not induce reliable synaptic potentiation in wild-type animals ($n = 9$ slices/9 mice), it evoked robust synaptic potentiation in transgenic slices ($n = 5$ slices/5 mice) (Fig. 3d). However, repetitive stimulation delivered at 5 Hz for 3 min (also giving 900 pulses) did not produce significant synaptic potentiation in either group of mice (transgenic, $n = 5$ slices/5 mice; mean $96.8 \pm 20.3\%$; wild-type, $n = 5$ slices/5 mice, mean $85.0 \pm 16.4\%$).

We then investigated long-lasting synaptic depression induced by low-stimulation frequency^{3,18}. Stimulation at 1 Hz produced similar LTD in both wild-type animals ($n = 6$ slices/6 mice, $76.0 \pm 9.3\%$) and transgenic mice ($n = 8$ slices/7 mice, $76.8 \pm 13.6\%$) (Fig. 3e). We also examined synaptic depression using another protocol in which low-frequency stimulation (5 Hz, 3 min) was applied 5 min after strong tetanic stimulation (100 Hz, 2×1 s)²⁰. Again, similar depression or depotentiation was induced in slices from transgenic mice ($n = 4$ slices/4 mice; $129.3 \pm 12.9\%$) and wild-type mice ($n = 6$ slices/6 mice, $111.1 \pm 14.8\%$). Figure 3f summarizes our results, showing that enhanced NMDA-receptor activation in transgenic mice results in selective enhancement of long-lasting synaptic potentiation evoked by stimulation of 10–100 Hz.

How might the selective enhancement of 10–100-Hz LTP responses contribute to learning and memory? As forebrain neurons often fire in this frequency range during behavioural experience (for example, hippocampal neurons fire at 4–12 Hz, known as the θ -rhythm, whereas various cortical neurons oscillate at 20–60 Hz, known as the γ -frequency), it is possible that the selective enhancement of potentiation by stimuli at ≥ 10 Hz in the transgenic mice is particularly meaningful. The conditional knockout of the NR1 subunit in the CA1 region leads to complete loss of synaptic changes

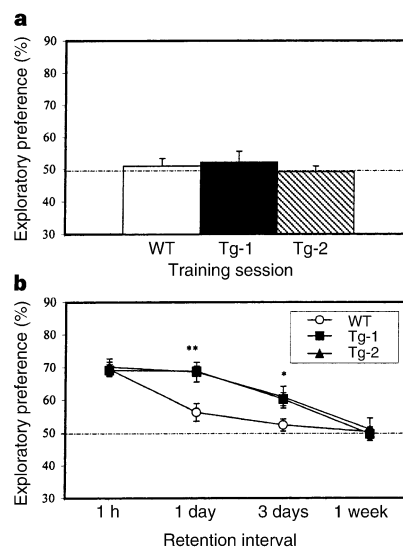


Figure 4 Enhanced novel-object recognition memory in transgenic mice. **a**, Exploratory preference in the training session. Dotted line represents performance at chance (50%). The amount of time spent exploring the two objects was the same for transgenic and wild-type mice. **b**, Enhanced exploratory preference in transgenic mice in retention test. Figure shows the temporary feature of the enhanced long-term memory in the transgenic mice (Tg-1, $n = 19$; Tg-2, $n = 8$; WT, $n = 14$). Data expressed as mean $\pm$ s.e.m. Single asterisk, $P < 0.05$; double asterisk, $P < 0.01$; *post hoc* analysis between transgenic and wild-type mice.

in the 1–100-Hz frequency range²¹ (Fig. 3f) and impairs performance in a spatial water maze. This indicates that a normal frequency response in this range is essential for learning and memory. A systematic downward shift (producing LTD) in this specific range can cause learning impairment¹⁵, whereas an upward shift (producing LTP) in all frequencies (1–100 Hz) is also deleterious to spatial learning²². These observations indicate the importance of normal 1–100-Hz responses for learning and memory.

To test whether the selective enhancement of responses to stimuli at 10–100 Hz represents an optimal plasticity curve, we conducted various learning tasks relevant to the forebrain regions. All behavioural experiments were performed with the experimenter blind to the genotype of each mouse.

We first used the novel-object-recognition task to measure visual recognition memory, which is evolutionarily conserved in species including humans and rodents and requires the hippocampus^{23–25}. To increase the difficulty of this task, we used a 5-min training protocol (see Methods). During training there was no significant difference in the amount of the time the mice spent exploring the two novel objects (as shown by the exploratory preference; Fig. 4a), indicating that both types of mouse have the same curiosity and motivation to explore the objects. During the retention test one of the familiar objects used in the training session was replaced with a third novel object, and animals were allowed to explore for 5 min. Both transgenic and wild-type mice exhibited similar preference towards the novel object at the 1-h retention test (Fig. 4b). This indicates that all groups were equally able to retain the memory of the old object for 1 hour. However, when retention tests were conducted either 1 day or 3 days later (Fig. 4b), both transgenic lines exhibited much stronger preference for the novel object than did the wild-type mice ($F(2, 38) = 5.448$, $P < 0.01$), indicating that transgenic mice have better long-term memory. A *post hoc* analysis using Dunnett's test reveals a significant difference between wild-type and either transgenic line at the 1-day ($P < 0.01$) or 3-day retention tests ($P < 0.01$), but not between the two transgenic lines. The enhanced long-term memory is therefore independent of the transgene integration locus. However, by 1 week after training, the

preference shown by transgenic mice also returned to the basal level.

We then assessed two forms of associative emotional memory in these mice: contextual and cued fear conditioning. Animals learn to fear either a neutral conditioned stimulus (such as tone) that has been paired with an aversive unconditioned stimulus (such as a foot shock) or a context in which the animals were conditioned by the pairing of conditioned and unconditioned stimuli. Contextual fear conditioning is hippocampus-dependent, whereas cued fear conditions is hippocampus-independent²⁶. Both these types of fear conditioning require the activation of NMDA receptors^{27,28}.

Both contextual and cued conditioning were measured at 1 h, 1 day and 10 days after training using separate groups of animals. We first analysed contextual fear memory. Transgenic mice consistently exhibited much stronger freezing responses (Fig. 5a–c). One-way analysis of variance (ANOVA) indicated no significant difference in immediate freezing between wild-type and transgenic mice, but a significant group difference in contextual freezing is found when they are tested at 1 h ($F(2, 24) = 6.062$, $P < 0.05$), 1 day ($F(2, 25) = 5.223$, $P < 0.05$), and 10 days ($F(1.18) = 4.576$, $P < 0.05$). Further *post hoc* analysis reveals the significant difference between the wild-type and either transgenic line ($P < 0.05$, respectively), but not between these two transgenic lines.

As the transgene is also abundantly expressed in the amygdala and the cortex, we then examined cued fear conditioning. One-way ANOVA indicated that freezing in response to the tone was also significantly elevated in transgenic mice compared to that in controls when tested at 1 h ($F(2, 24) = 4.672$, $P < 0.05$), 1 day ($F(2, 25) = 5.518$, $P < 0.01$) or 10 days ($F(1.18) = 6.498$, $P < 0.05$) after training (Fig. 5d–f). A *post hoc* analysis shows the significant difference between wild-type and either transgenic line ($P < 0.05$, respectively). The enhanced contextual and cued fear memories in transgenic mice were not due to altered nociceptive responses, as the minimal amount of current required to elicit three stereotypical behaviours (flinching/running, jumping and vocalizing) were similar in wild-type and transgenic mice (data not shown).

We then conducted two experiments to measure emotional learning using the fear-extinction paradigm²⁹. If animals are repeatedly exposed to the context or the conditioned stimulus (tone) without the unconditioned stimulus (shock), the context or conditioned stimulus will lose its ability to produce fear responses. This reduction in conditioned fear is known as fear extinction and is NMDA-dependent²⁹. It is thought to involve the formation of a new

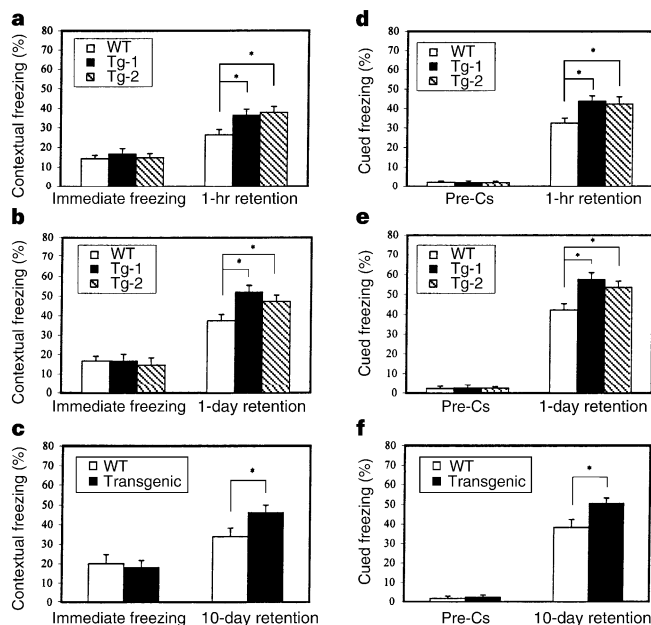


Figure 5 Enhancement of both contextual and cued fear memory in transgenic mice. **a–c**, Contextual conditioning 1 h, 1 day and 10 days after training, respectively. **d–f**, Cued fear conditioning 1 h, 1 day and 10 days after training, respectively. Each point represents data collected from 8–10 mice per group (WT, Tg-1 or Tg-2). The value in each column represents percentage of freezing rate and is expressed as mean $\pm$ s.e.m. Asterisk, $P < 0.05$, *post hoc* analysis between wild-type and transgenic mice.

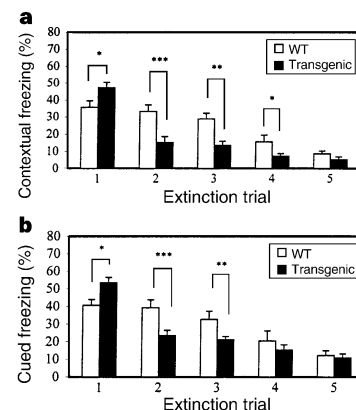


Figure 6 Transgenic mice exhibit faster learning in fear extinction. **a**, Faster fear extinction to contextual environment in transgenic mice. Either wild-type ($n = 8$) or transgenic (Tg-1, $n = 7$; Tg-2, $n = 8$; data plotted together) mice were given the same single CS/US pairing training as described in Fig. 5 and were then subjected to five extinction trials 24 h after training. **b**, Faster fear extinction to the tone in transgenic mice. The value in each column represents percentage of freezing rate; data are expressed as mean $\pm$ s.e.m. Asterisk, $P < 0.05$; double asterisk, $P < 0.01$; triple asterisk, $P < 0.001$, *post hoc* analysis.

memory rather than the passive decay or erasure of the original memory, because the original associations remain intact following extinction²⁹. We examined fear extinction using a 5-extinction-trial protocol. When we measured the initial fear response 24 h after training, transgenic mice again showed much stronger fear responses than controls in both contextual and cued fear conditioning (Fig. 6). However, transgenic mice exhibited much less freezing during subsequent exposure to either the context or the tone than did wild-type mice (Fig. 6a, b). A two-way ANOVA indicated that, although both wild-type and transgenic mice decreased their freezing responses to contextual ($F(4, 80) = 86.247$, $P < 0.001$) or cued extinction ($F(4, 80) = 78.415$, $P < 0.001$), there was a significant group difference between transgenic and wild-type mice in contextual extinction ($F(2, 20) = 8.595$, $P < 0.01$) and in cued extinction ($F(2, 20) = 7.778$, $P < 0.01$). A *post hoc* analysis revealed a significant difference in the freezing responses between wild-type and transgenic mice at the second or third extinction trial in both contextual conditioning ($P < 0.05$) and cued conditioning ($P < 0.05$). Similar faster fear extinction was also observed if the experiments were conducted 1 h after fear conditioning (data not shown). Therefore, transgenic mice are quicker to learn to disassociate the previously paired events.

Finally, we tested spatial learning in transgenic mice using the hidden-platform water maze, which requires the activation of NMDA receptors in the hippocampus^{21,30}. As shown in Fig. 7, the latency to escape to the platform in both wild-type and transgenic mice decreased following the training sessions. However, there was a significant group difference throughout sessions ($F(1, 26) = 9.655$, $P < 0.01$), indicating that spatial learning in transgenic mice was faster than in wild-type mice. Moreover, a *post hoc* analysis reveals a significant difference at the third session ($P < 0.05$), confirming better learning in transgenic mice. In addition, enhanced spatial learning in transgenic mice was also evident in the first transfer test conducted after the third training session. In comparison with controls, transgenic mice already exhibited a clear preference for

the targeted quadrant in which the platform was previously located ($P < 0.05$; Student's *t*-tests) (Fig. 7b). With additional training, control mice showed the same preference as transgenic mice, as measured by either escape latency or place preference in the second transfer test after the last (sixth) session. Thus, the transgenic mice perform better than their wild-type littermates in this spatial task.

We have shown that the NMDA receptor does serve as a graded molecular switch for gating the age-dependent threshold for synaptic plasticity and memory formation, thus substantially validating Hebb's learning rule. This further demonstrates that NMDA-dependent modifications of synaptic efficacy represent the unifying mechanism underlying a variety of associative learning and memory. Our data further indicate that neural activities at the 10–100 Hz range in the forebrain may be crucial for coding and storage of learned information. In addition, the identification of NR2B as a molecular switch in the memory process has indicated a potential new target for the treatment of learning and memory disorders. This study also reveals a promising strategy for the creation of other genetically modified mammals with enhanced intelligence and memory. □

Methods

For extended methods, see Supplementary Information.

Production and basic characterization of transgenic mice.

The transgenic founders were produced by pronuclear injection of the linearized DNA into C57B/6 inbred zygotes as described¹⁶, and then intercrossed with B6/CBF1 for various analyses. F2 wild-type mice on this hybrid background consistently showed excellent learning. This mating strategy, therefore, sets a high standard for our behavioural enhancement experiments. For detailed procedures for genotyping, northern blot, western blot and *in situ* hybridization, see Supplementary Information.

Hippocampal cell culture and recording.

Primary cultures of hippocampal neurons were prepared from individual neonatal mice (P1). Whole-cell patch recordings were carried out as described¹⁷. Recordings were made with a 200B integrating patch-clamp amplifier (Axon Instruments) with a 1 kHz (8-pole Bessel) low-pass filter. Data were digitized at 10 kHz using a Digidata 1200B A/D converter (Axon Instruments). Glutamate currents were evoked by iontophoresis as described¹⁷. Briefly, following a one minute incubation in the FM 1-43 solution, neurons, continuously perfused with tyrode, were visualized under confocal microscopy. Following placement of the iontophoresis electrode, brief (1 ms) glutamate pulses of varied amplitudes were delivered to an isolated FM-labelled presynaptic bouton.

Hippocampal slice recording.

Transverse slices of the hippocampus from transgenic and wild-type littermates (4–6-months old) were rapidly prepared and maintained in an interface chamber. A bipolar tungsten stimulating electrode was placed in the stratum radiatum in the CA1 region and extracellular field potentials were recorded using a glass microelectrode (3–12 MΩ, filled with ACSF) also in the stratum radiatum. Test responses were elicited at 0.02 Hz. We also measured PPF of the response at various interpulse intervals (25–400 ms). Data are presented as mean $\pm$ s.e.m. One-way ANOVA (with Duncan's multiple range test for *post hoc* comparison) and Student's *t*-test were used for statistical analysis.

Behavioural tests.

Adult transgenic and wild-type mice (3–6-month-old littermates) were used throughout all behavioural tests. Unless stated otherwise, ANOVA and *post hoc* Dunnett's test were used to determine genotype effects on the behavioural responses.

Novel object recognition task.

Mice were individually habituated to an open-field box (20 $\times$ 20 $\times$ 10 high inches) for 3 days. During training sessions, two novel objects were placed into the open field and the animal was allowed to explore for 5 min. The time spent exploring each object was recorded. During retention tests, the animals were placed back into the same box, in which one of the familiar objects used during training was replaced by a novel object, and allowed to explore freely for 5 min. A preference index, a ratio of the amount of time spent exploring any one of the two objects (training session) or the novel one (retention session) over the total time spent exploring both objects, was used to measure recognition memory.

Fear conditioning task.

We used a fear conditioning shock chamber (10 $\times$ 10 $\times$ 15 inches high) and TruScan multi-parameter activity monitors (Coulbourn Instruments). The conditioned stimulus (CS) used was an 85dB sound at 2,800 Hz, and the unconditioned stimulus (US) was a continuous scrambled foot shock at 0.75 mA. After the CS/US pairing, the mice were allowed to stay in the chamber for another 30 s for measurement of immediate freezing. During the retention test, each mouse was placed back into the shock chamber and the

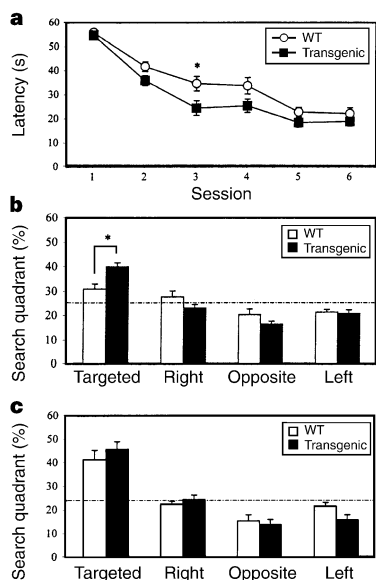


Figure 7 Enhanced performance in the hidden-platform water maze task by transgenic mice. **a**, Escape latency (mean $\pm$ s.e.m.) in water maze training (Tg-1, $n = 13$; wild-type, $n = 15$). **b**, Place preference in the first transfer test conducted at the end of the third training session. Transgenic mice spent more time in the target quadrant than other quadrants, whereas control mice did not show any preference for the target quadrant at this stage. **c**, Place preference in the second transfer test carried out at the end of the sixth training session. Both transgenic and wild-type mice exhibited strong preference for the target quadrant where the hidden platform was previously located. Asterisk, $P < 0.05$, *post hoc* analysis in **a**, and Student's *t*-test in **b**, between transgenic mice and controls.

freezing response was recorded for 3 min (contextual conditioning). Subsequently, the mice were put into a novel chamber and monitored for 3 min before the onset of the tone (pre-CS). Immediately after that, a tone identical to the CS was delivered for 3 min and freezing responses were recorded (cued conditioning).

Fear extinction experiment.

Twenty-four hours after training, the mice were given the first extinction trial. Each extinction trial consisted of contextual and cued extinction. The mice were first put individually into the shock chamber and observed for 3 min in the absence of the US to measure contextual extinction. Then, the mice were transferred into a novel box for measurement of the pre-CS response for 3 min (in the absence of the tone), and subsequently of cued conditioning for 3 min in the presence of the training tone. Following this, the same four extinction trials were given at 2-h intervals and freezing responses were recorded.

Water maze task.

The water-maze apparatus is a circular pool (1.2 m in diameter). The procedure was essentially as described²¹. The training protocol consisted of six sessions (4 trials per session per day). The navigation of the mice was tracked by a videocamera, and the escape latency to the platform was recorded. In addition, we carried out two transfer tests. The first was done at the end of the third session and the second at the end of the last session. During the transfer test, the platform was removed and the mice were allowed to swim in the pool for 60 s. The time spent in each quadrant was recorded. Student's *t*-test was used to determine genotype effect on the spatial preference.

Received 9 February; accepted 14 June 1999.

1. Stevens, C. F. & Sullivan, J. Synaptic plasticity. *Curr. Biol.* **8**, R151–153 (1998).
2. Bliss, T. V. & Collingridge, G. L. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* **361**, 31–39 (1993).
3. Bear, M. F. & Malenka, R. C. Synaptic plasticity: LTP and LTD. *Curr. Opin. Neurobiol.* **4**, 389–399 (1994).
4. Bourne, H. R. & Nicoll, R. Molecular machines integrate coincident synaptic signals. *Cell (suppl.)* **72**, 65–75 (1993).
5. Nakanishi, S. Molecular diversity of glutamate receptors and implications for brain function. *Science* **258**, 597–603 (1992).
6. Hollmann, M. & Heinemann, S. Cloned glutamate receptors. *Annu. Rev. Neurosci.* **17**, 31–108 (1994).
7. Monyer, H. *et al.* Heteromeric NMDA receptors: molecular and functional distinction of subtypes. *Science* **256**, 1217–1221 (1992).
8. Monyer, H., Burnashev, N., Laurie, D. J., Sakmann, B. & Seeburg, P. H. Developmental and regional expression in the rat brain and functional properties of four NMDA receptors. *Neuron* **12**, 529–540 (1994).
9. Sheng, M., Cummings, J., Roldan, L. A., Yan, Y. N. & Yan, L. Y. Changing subunit composition of heteromeric NMDA receptors during development of rat cortex. *Nature* **368**, 144–147 (1994).
10. Okabe, S. *et al.* Hippocampal synaptic plasticity in mice overexpressing an embryonic subunit of the NMDA receptor. *J. Neurosci.* **18**, 4177–4188 (1998).
11. Carmignoto, G. & Vicini, S. Activity-dependent decrease in NMDA receptor responses during development of the visual cortex. *Science* **258**, 1007–1011 (1992).
12. Hestrin, S. Activation and desensitization of glutamate-activated channels mediating fast excitatory synaptic currents in the visual cortex. *Neuron* **9**, 991–999 (1992).
13. Kuhl, P. K. Learning and representation in speech and language. *Curr. Opin. Neurobiol.* **4**, 812–822 (1994).
14. Konishi, M. Birdsong for neurobiologists. *Neuron* **3**, 541–549 (1989).
15. Mayford, M., Wang, J., Kandel, E. R. & O'Dell, T. J. CaMKII regulates the frequency-response function of hippocampal synapses for the production of both LTD and LTP. *Cell* **81**, 891–904 (1995).
16. Tsien, J. Z. *et al.* Subregion- and cell type-restricted gene knockout in mouse brain. *Cell* **87**, 1317–1326 (1996).
17. Liu, G., Choi, S. & Tsien, R. W. Variability of neurotransmitter concentration and nonsaturation of postsynaptic AMPA receptors of synapses in hippocampal cultures and slices. *Neuron* **22**, 395–409 (1999).
18. Dudek, S. M. & Bear, B. F. Bidirectional long-term modification of synaptic effectiveness in the adult and immature hippocampus. *J. Neurosci.* **13**, 2910–2918 (1993).
19. Harris, K. M. & Teyler, T. J. Developmental onset of long-term potentiation in area CA1 of the rat hippocampus. *J. Physiol. (Lond.)* **346**, 27–48 (1984).
20. Stubli, U. & Chun, D. J. Factors regulating the reversibility of long-term potentiation. *J. Neurosci.* **16**, 853–860 (1996).
21. Tsien, J. Z., Huerta, P. T. & Tonegawa, S. The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory. *Cell* **87**, 1327–1338 (1996).
22. Migaud, M. *et al.* Enhanced long-term potentiation and impaired learning in mice with mutant postsynaptic density-95 protein. *Nature* **396**, 433–439 (1998).
23. Reed, J. M. & Squire, L. R. Impaired recognition memory in patients with lesions limited to the hippocampal formation. *Behav. Neurosci.* **111**, 667–675 (1997).
24. Myhrer, T. Exploratory behavior and reaction to novelty in rats with hippocampal perforant path systems disrupted. *Behav. Neurosci.* **102**, 356–362 (1988).
25. Mumby, D. G. *et al.* Ischemia-induced object-recognition defects in rats are attenuated by hippocampal ablation before or soon after ischemia. *Behav. Neurosci.* **110**, 266–281 (1996).
26. Phillips, R. G. & LeDoux, J. E. Differential contribution of amygdala and hippocampus to cued and contextual fear conditioning. *Behav. Neurosci.* **106**, 274–285 (1992).
27. Kim, J. J., Fanselow, M. S., DeCola, J. P. & Landeira-Fernandez, J. Selective impairment of long-term but not short-term conditional fear by the *N*-methyl-D-aspartate antagonist APV. *Behav. Neurosci.* **106**, 591–596 (1992).
28. Davis, M., Hitchcock, J. & Rosen, J. B. In *The Psychology of Learning and Memory* (ed. Bower, G. H.) (Academic, New York, 1987).
29. Falls, W. A., Miserendino, M. J. D. & Davis, M. Extinction of fear-potentiated startle: blockade by infusion of an NMDA antagonist into the amygdala. *J. Neurosci.* **12**, 854–863 (1992).
30. Morris, R. G., Garrud, P., Rawlins, J. N. & O'Keefe, J. Place navigation impaired in rats with hippocampal lesions. *Nature* **24**, 681–683 (1982).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank G. Bracket and L. Antonucci and their staff for the maintenance of Princeton Mouse Facility, M. Mayford and E. Kandel for providing the CaMKII promoter, S. Tonegawa for support of J.R.T.'s early cloning work (1995) on plasmid p279NB from which pJT-NR2B is derived and D. Prout for secretarial assistance and proofreading. This work was supported by Princeton University, Beckman Foundation and NIH (J.Z.T.).

Correspondence and requests for materials should be addressed to J.Z.T. (e-mail: jsien@molbio.princeton.edu).

Growth-cone attraction to netrin-1 is converted to repulsion by laminin-1

Veit H. Höpker*[§], Derryck Shewan†[§], Marc Tessier-Lavigne‡, Mu-ming Poo* & Christine Holt†

* Department of Biology, University of California at San Diego, La Jolla, California 92093-0357, USA

† Department of Anatomy, University of Cambridge, Cambridge CB2 3DY, UK
‡ Howard Hughes Medical Institute, Departments of Anatomy, and Biochemistry and Biophysics, University of California, San Francisco, California 94143-0452, USA

§ These authors contributed equally to this work.

Growing axons are guided by both diffusible and substrate-bound factors^{1–3}. Growth cones of retinal neurons exhibit chemo-attractive turning towards the diffusible factor netrin-1 *in vitro*⁴ and are guided into the optic nerve head (ONH) by localized netrin-1 (ref. 5). Here we report that, in *Xenopus*, laminin-1 from the extracellular matrix (ECM), converts netrin-mediated attraction into repulsion. A soluble peptide fragment of laminin-1 (YIGSR) mimics this laminin-induced conversion. Low levels of cyclic AMP in growth cones also lead to the conversion of netrin-induced attraction into repulsion⁶, and we show that the amount of cAMP decreases in the presence of laminin-1 or YIGSR, suggesting a possible mechanism for laminin's effect. At the netrin-1-rich ONH, where axons turn sharply to leave the eye, laminin-1 is confined to the retinal surface. Repulsion from the region in which laminin and netrin are coexpressed may help to drive axons into the region where only netrin is present, providing a mechanism for their escape from the retinal surface. Consistent with this idea, YIGSR peptides applied to the developing retina cause axons to be misdirected at the ONH. These findings indicate that ECM molecules not only promote axon outgrowth, but also modify the behaviour of growth cones in response to diffusible guidance cues.

The finding that growth cones rapidly turn away from netrin-1 when little cytosolic cAMP is present⁶ suggested that coincident signals impinging on the same intracellular signalling pathway can influence the behaviour of growth cones *in vivo* as they navigate complex terrains. Because laminin-1 is abundant in the developing optic pathway and can promote axon outgrowth from retinal neurons^{7–9}, we investigated whether it can act as a coincident regulator of netrin-1-guided growth. When grown on glass, poly-D-lysine or fibronectin, *Xenopus* retinal growth cones consistently showed attractive turning in the presence of a netrin-1 gradient (Fig. 1a–d), mediated by the receptor Deleted in colorectal cancer (DCC)⁴. In contrast, when plated on poly-D-lysine coated with a low concentration of laminin-1 (1 µg ml^{–1}), retinal axons showed heterogeneous responses to the netrin-1 gradient, composed of either

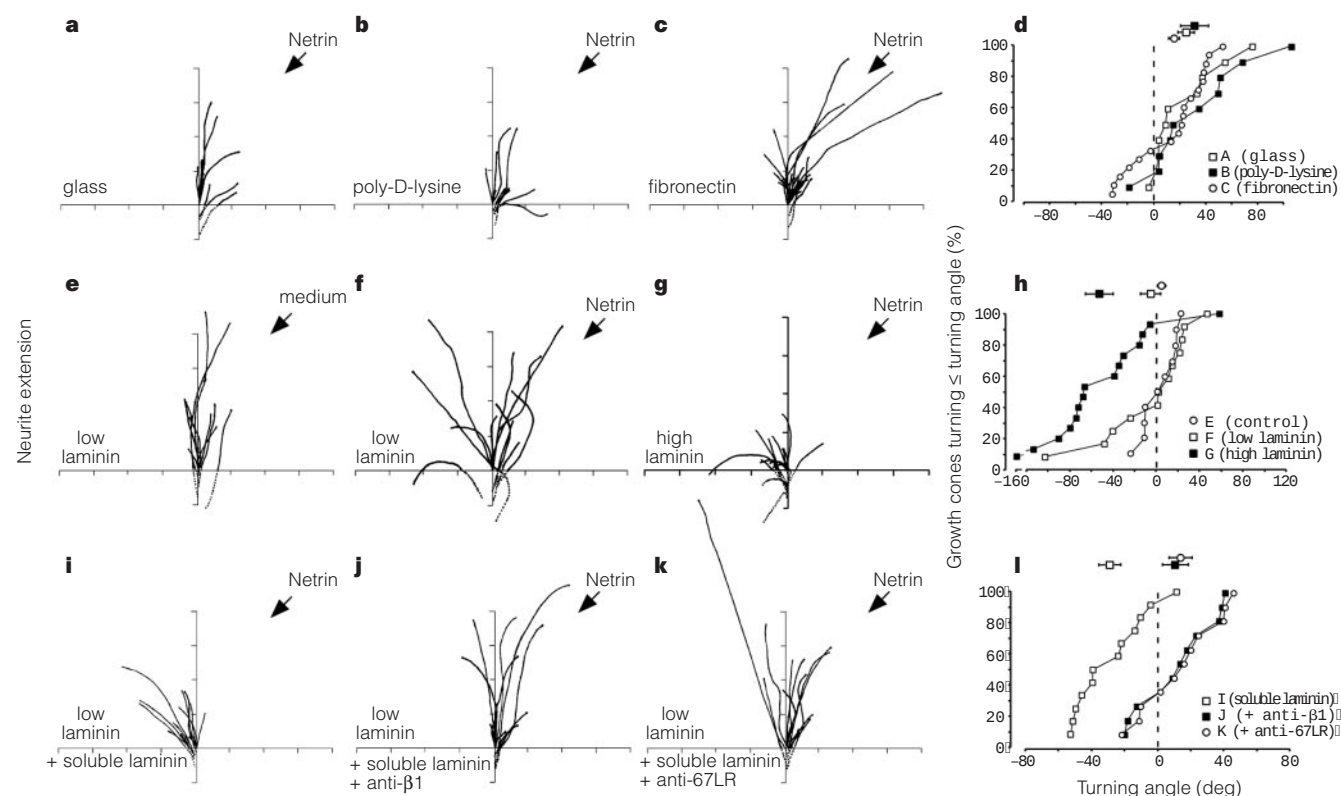


Figure 1 Laminin-1 converts netrin-1-induced growth-cone attraction into repulsion. Superimposed traces of neurite trajectories (**a–c**, **e–g**, **i–k**) and cumulative plots of turning angles (**d**, **h**, **l**) from growth-cone turning assays. Arrows indicate direction of gradient^{4,6}. Growth cones cultured on glass (**a**), poly-D-lysine (**b**) and fibronectin (**c**) turn towards netrin-1. On low laminin-1 ($1 \mu\text{g ml}^{-1}$), no turning bias occurs with control medium in the pipette (**e**), whereas a mix of attractive and repulsive turning occurs in a

netrin-1 gradient (**f**). On high laminin-1 ($20 \mu\text{g ml}^{-1}$), growth cones exhibit strong repulsion (**g**, **h**). Adding soluble laminin ($10 \mu\text{g ml}^{-1}$), causes marked repulsion (**i**, **l**) that is blocked by antibodies to β1-integrin or to 67LR (**j–l**). Reagents added to culture medium at the start of the assay are shown below abscissa in **i–k**. Isolated symbols represent mean values. Trajectory plot axes are divided into 10-μm intervals.

attractive or repulsive responses, with an average turning angle close to 0° (Fig. 1f, h). In control experiments, the average turning angle was also close to 0° but neurite trajectories were relatively straight, without bias to or from the netrin-ejecting pipette (Fig. 1f, h). When the substrate was precoated with a high concentration of laminin-1 ($20 \mu\text{g ml}^{-1}$), nearly all growth cones exhibited repulsive turning responses (Fig. 1g, h). A marked repulsion of all retinal growth cones in the netrin-1 gradient was also observed when soluble laminin-1 ($10 \mu\text{g ml}^{-1}$) was added before we positioned the pipette (Fig. 1i, l). The αβ1 integrin is a cell-surface laminin-1 receptor that mediates the effects of laminin-1 in promoting growth of retinal axons during development^{10,11}, so we investigated whether integrins also mediate the laminin-induced conversion. Addition of a β1 integrin function-blocking antibody¹² abolished the repulsive responses to netrin-1 in soluble laminin-1, and attractive turning similar to that on a fibronectin substrate occurred (Fig. 1j, l). Thus, the conversion of the netrin-1 turning response seen with high laminin-1 may be mediated by β1 integrins.

To understand the conversion of the netrin-1-induced turning response to repulsion, we examined the effect of two laminin-1 fragments: YIGSR, a five-amino-acid peptide in the laminin-1 B1 chain, which mediates cell attachment but has no neurite growth-promoting activity¹³; and a 19-amino-acid peptide containing the IKVAV sequence from the laminin-1 A1 chain, which promotes neurite growth and stimulates the formation of neuritic processes¹⁴. When the YIGSR peptide was added to the culture medium, retinal axons showed a complete switch to repulsive turning (Fig. 2a). The YIGSR effect was dose dependent, with turning neutralized at $1 \mu\text{g ml}^{-1}$ and complete conversion to repulsion at $10 \mu\text{g ml}^{-1}$ (Fig. 2b). A synthetic peptide with the reverse order of amino acids, RSGIY, and the IKVAV-containing peptide, had no effect on

the turning response (Fig. 2a). A laminin-1 receptor of relative molecular mass 67,000 (67LR), which binds YIGSR¹⁵, is expressed in the developing retina¹⁵ and may stabilize laminin-1 binding to integrin receptors^{16,17}. In the presence of a function-blocking antibody directed against 67LR¹⁸, the ability of both laminin-1 and YIGSR to switch the direction of growth-cone turning was abolished (Figs 1k, l and 2b). Thus YIGSR may be involved in mediating the conversion of the turning response.

To test whether the turning behaviour of a growth cone can readily be modulated between attraction and repulsion in the same neuron, we performed sequential turning experiments on single growth cones (Fig. 3). An attractive response was first induced over 1 h in a netrin-1 gradient (Fig. 3a, b). The pipette was repositioned at 45° to the new direction of growth, the YIGSR peptide was added and the same netrin-1 gradient applied for 1 h more (Fig. 3c, d), in which a repulsive response was observed. Similar reversals of turning responses were observed in five other growth cones (Fig. 3e, f).

The nature of laminin-1 signalling in these neurons is not known. Previous studies have shown that reducing the amount of intracellular cAMP or protein kinase A activity can convert the turning response of *Xenopus* spinal neurons to brain-derived neurotrophic factor (BDNF) from attraction to repulsion^{6,19}. Bath application of Sp-cAMPS (a cAMP analogue that activates PKA²⁰) reversed the effect of the YIGSR peptide with most growth cones showing attractive turning towards netrin-1 (Fig. 2a). The partial repulsive turning observed on low laminin-1 (see Fig. 1f) was converted to complete repulsion by adding Rp-cAMPS, a non-hydrolysable cAMP analogue that competes with cAMP²⁰ (average turning angle decreased from -5.9° to -16.7°). Increasing cytosolic cAMP reportedly blocks laminin-1-induced outgrowth in goldfish retinal neurons²¹, and the β1 integrin-mediated migration of T cells to

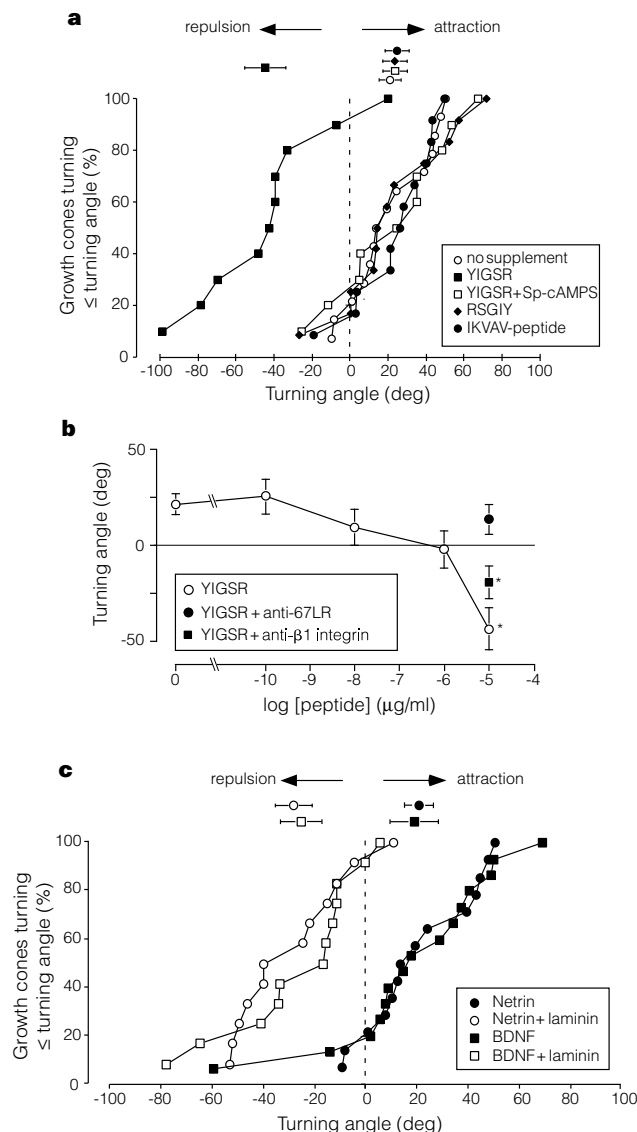


Figure 2 YIGSR mimics laminin-1 and causes repulsive turning in netrin-1 gradients. **a**, Cumulative distribution of turning angles of retinal neurites grown on poly-D-lysine in a netrin-1 gradient in the presence of bath applied laminin-1 peptides and Sp-cAMPS. **b**, Dose-response plot (mean $\pm$ s.e.) of YIGSR and the effects of blocking antibodies ($n = 10-13$; * $P < 0.05$, Kolmogorov-Smirnov test). The difference between the data for YIGSR and YIGSR + anti-67 LR was significant, but that between YIGSR and YIGSR + anti- $\beta 1$ integrin was not. **c**, Attractive turning in gradients of both netrin-1 or BDNF is reversed by laminin-1. Explants grown on poly-D-lysine or low laminin-1 in combination with soluble laminin-1. In both cases, laminin-1-induced conversion of turning angles was significantly different from cultures grown on poly-D-lysine ($P < 0.05$).

fibronectin depends on a $\beta 1$ integrin-associated G protein and cAMP²². Consistent with the idea that the laminin-1 effect is related to cAMP modulation, we found that the attractive responses of retinal growth cones in a BDNF gradient were also converted to repulsive responses by plating the cells on laminin-1-coated substrates (Fig. 2c).

To test whether laminin-1 and YIGSR alter the amount of cAMP within growth cones, we used an antibody that binds to free cAMP in the cytosol^{23,24}. cAMP intensity increased in RGC growth cones when forskolin, a specific activator of adenylyl cyclase, was added (Fig. 4). Netrin-1 alone significantly increased cAMP levels but, when added with laminin-1 or YIGSR, cAMP decreased, indicating that laminin-1 and YIGSR may prevent netrin-1-induced increases in cAMP (Fig. 4). Laminin-1 and YIGSR alone also lowered cAMP (data not shown).

These findings indicate that netrin-1 and laminin-1 have a

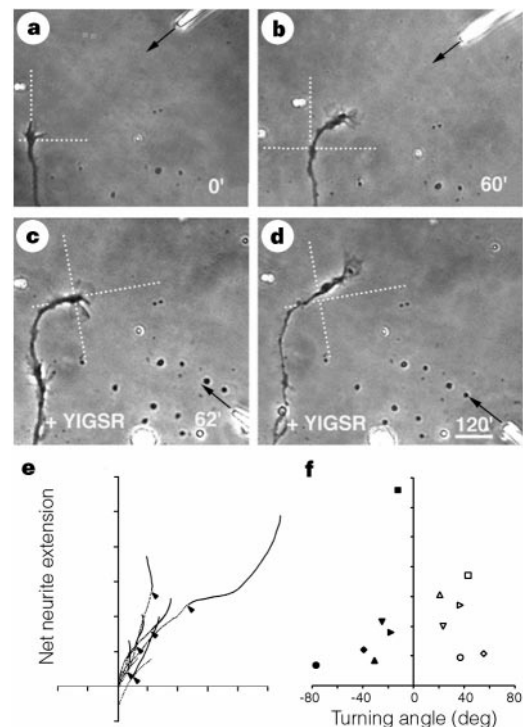


Figure 3 Opposite netrin-1-induced turning behaviours of single growth cones with and without YIGSR peptide. **a**, Netrin-1 gradient applied to a retinal growth cone. **b**, After 1 h, the growth cone turned towards the netrin-1 source. **c**, The pipette was repositioned and YIGSR supplemented to the medium. **d**, After a further 1 h, the same growth cone turned away from netrin-1. Arrows indicate the direction of netrin-1 gradient. Scale bar, 20 μm . **e**, Superimposed neurite trajectories of six growth cones over 2 h. Arrowheads indicate the position of the growth cones after 1 h. Trajectories during the first hour are dashed lines, those in second hour are solid lines. **f**, Scatter plot of turning angles and neurite extension after 1 h (open symbols) and 2 h (filled symbols). In **e** and **f**, unlabelled tick marks represent 10 μm .

combinatorial role in guiding retinal axons. Perturbation of $\beta 1$ integrin function impairs RGC axon growth in the retina¹¹, suggesting that ECM molecules are necessary for promoting axon growth out of the eye. Immunostaining sections of embryonic retina showed that laminin-1 is localized to the vitreal surface of the retina along the entire surface of the optic-fibre layer where RGC axons grow. Near the ONH, laminin-1 is expressed only at the entrance where growth cones turn away from the vitreal surface to leave the retina²⁵, whereas netrin-1 is expressed throughout the ONH^{4,5} (Fig. 5a, b). Misrouting defects occur at the ONH in mutant netrin-1 and DCC mice⁵, showing that netrin-1 is needed for the ONH to form appropriately. To test the idea that laminin-1 is also important for steering retinal growth cones at the ONH, YIGSR or a control peptide was applied to the developing retina when the axons first extend across its surface. The pattern of axon growth was assayed using anti-acetylated tubulin immunostaining in retinal wholemounts. In control eyes, axons grew towards the ONH, where they converged to form a single, thick fascicle that left the eye (Fig. 5c). In YIGSR-treated eyes, axons often failed to converge into a single fascicle at the ONH, instead forming several widely separated fascicles (Fig. 5d), some of which did not appear to leave the eye, and axon trajectories close to the ONH were often disorganized (Fig. 5e). The YIGSR-induced defects were dose dependent (data not shown).

Our results show that ECM molecules not only function in substrate-mediated adhesion, but also modulate the response of growing axons to secreted guidance cues. Increasing cAMP on one side of the growth cone has previously been shown to induce attractive turning, and a cAMP gradient of just 10% may be required²⁶. At the entrance to the ONH, distribution of cAMP

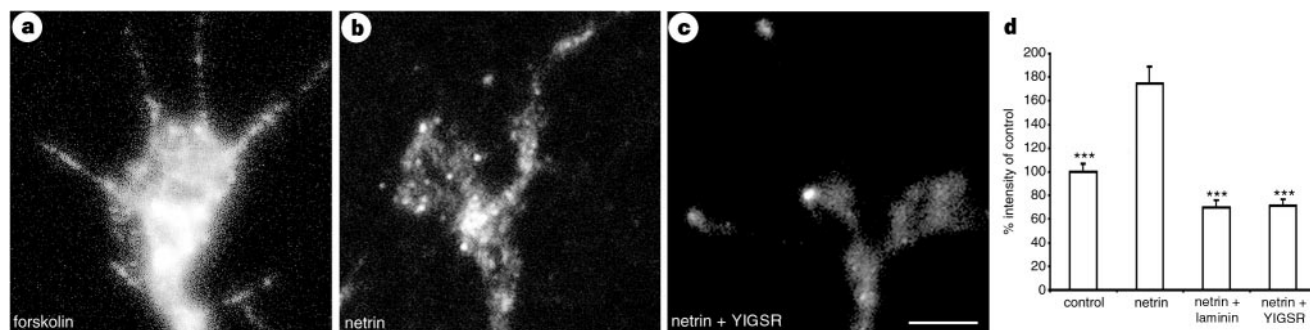


Figure 4 Laminin-1 and YIGSR reduce cAMP levels in retinal growth cones. Growth cones immunostained with anti-cAMP antibody exhibit intense staining after incubation with forskolin (**a**) and medium intensity with netrin-1 alone (**b**). cAMP staining is significantly diminished with simultaneous addition of YIGSR and netrin-1 (**c, d**) and laminin-1 plus netrin-1 (**d**). cAMP intensity measurements were normalized to CFDA (Molecular Probes; 20 μ M, 15 min) fluorescence in growth cones in similar conditions. In **d**, values are the

mean $\pm$ s.e. for 5 experiments. Total number of cAMP stained growth cones analysed: control, no reagents added to medium, 53; netrin, 62; netrin + laminin, 41; YIGSR + netrin, 46. The value for growth cones treated with 20 μ M forskolin was 367 ± 120 ($P < 0.01$, $n = 33$; not shown). Laminin-1 and YIGSR alone also lowered cAMP below the mean netrin value (data not shown). Scale bar, 4 μ m, *** $P < 0.001$.

modulators could establish a cAMP gradient across the growth cone (low on the side with laminin and netrin, high on the netrin-only side), causing turning away from the retinal surface. Changes in the amount of cAMP probably regulate, rather than directly bring about, turning. Netrin-1 has been proposed to activate polymeriza-

tion of cytoskeletal proteins when protein kinase A is active and depolymerization when it is inactive²⁷. Thus, the netrin-1-induced increase in cAMP could help to promote a polymerization-based attractive response, and laminin-1 could promote a repulsive response by favouring netrin-1-mediated depolymerization. Laminin-1 is expressed widely in the developing brain and is localized to the outer pial surface where pioneering axon tracts commonly form. Modulation by laminin-1 might be particularly important where axons leave the ECM-rich surface and dive deep into the central nervous system to form interior pathways. □

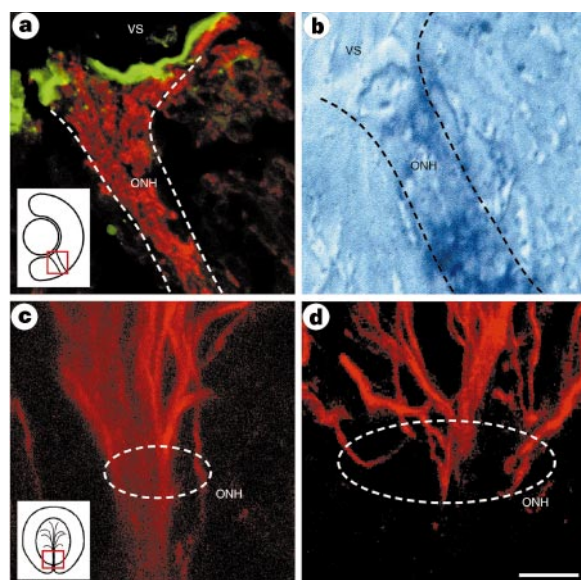


Figure 5 Exogenous YIGSR induces defects at the ONH where endogenous distribution of laminin-1 and netrin-1 is spatially discrete. **a**, Transverse section of the ONH (stage 37/38) double labelled with anti-acetylated tubulin (axons, red) and anti-laminin-1 (green). Laminin-1 is localized to the vitreal surface (VS) of the ONH and to the optic-fibre layer in the rest of the retina. **b**, Netrin-1 expression (*in situ* hybridization⁴) is localized to and throughout the ONH. **c**, View of retinal surface of control wholemount retina showing anti-acetylated-tubulin-positive axon fascicles converging into a single narrow bundle at the ONH (dashed ellipse). **d**, YIGSR-treated retina exhibits disorganized axon bundling at the ONH; several widely spread fascicles are apparent, some showing aberrant trajectories. The reverse sequence peptide (RSGIY) had no significant effect compared with controls (**e**). Dorsal is up in **c** and **d**. ** $P < 0.01$. Scale bar, 50 μ m.

Methods

Reagents.

The YIGSR peptide (laminin B1 chain 929–933) is Tyr-Ile-Gly-Ser-Arg (Sigma). The reverse order peptide, RSGIY, was a gift from H. K. Kleinman. The IKVAV-containing peptide (Cys–laminin A chain 2091–2108) is Cys-Ser-Arg-Ala-Arg-Lys-Gln-Ala-Ala-Ser-Ile-Lys-Val-Ala-Val-Ser-Ala-Asp-Arg (Sigma). All peptides were at a final concentration of 10 μ g ml⁻¹ for turning assays and 50 μ g ml⁻¹ for exposed eye preparations and cAMP analysis. The cAMP analogues, Sp-cAMPS and Rp-cAMPS (Calbiochem), were added at final concentrations of 20 μ M. For culture substrates, glass coverslips (BDH) were coated with 100 μ g ml⁻¹ of poly-D-lysine (Sigma), followed by fibronectin (1 μ g ml⁻¹; Sigma) or laminin-1 (1 μ g ml⁻¹ or 20 μ g ml⁻¹; Sigma). Soluble laminin-1 was added at 10 μ g ml⁻¹.

For antibody blocking experiments, cultures were preincubated with anti-67LR anti-serum (1:10, from H. K. Kleinman) or anti- β 1 integrin antibody (3 mg ml⁻¹, from K. Yamada) for 30 min before the turning assay. These antibodies did not affect the rate of extension of retinal neurites. For immunolabelling, the following antibodies were used: anti-acetylated tubulin (clone 6-11B-1; Sigma); anti-laminin-1 (gift from J. Fawcett); anti-cAMP (gift from T. Wiemelt); and biotin- (Vector), Cy3- and streptavidin-fluorescein-conjugated (Jackson) secondary antibodies. All were used at 1:100 dilutions.

Growth-cone turning assay.

Retinal explants from stage 23/24 *Xenopus*²⁸ eye primordia were prepared as described^{4,29} and plated on poly-D-lysine, low laminin-1 or fibronectin. Retinal explants were cultured for 18 h at room temperature before the turning assays. Netrin-1 gradients were produced as described²⁶ by pulsatile ejection of a concentrated solution of purified recombinant chick netrin-1 (5 μ g ml⁻¹) from a glass micropipette by using a pressure application system (Picospritzer, General Valve). The pipette tip was placed 100 μ m from the growth cone, at 45° to the direction of growth, resulting in a $\sim 10^3$ -fold dilution of netrin-1 concentration from pipette to growth cone²⁶.

Images were captured by a CCD camera (Sony XC-75) using a phase-contrast inverted microscope (Nikon Diaphot). Neurite extension over 1 h was determined by measuring the neurite trajectory using a digitizer. The turning angle was defined as the angle between the original direction of growth and a line drawn between the positions of the growth cone before and after 1 h. Only growth cones that extended > 5 μ m were included in the analysis²⁶. Extension rates were not altered by netrin-1. Net neurite extension increased on laminin-1 substrates consistent with the neurite promoting laminin-1 action⁷, and assay times in high laminin-1 were shortened to 20–30 min because neurite shafts detached over longer time periods. BDNF (Promega) was used at 50 μ g ml⁻¹. Statistical analysis used the GBSTAT software package.

cAMP analysis.

Stage 24 *Xenopus* retinal explants were plated overnight on low laminin-1. Netrin-1 (300 ng ml⁻¹) was bath applied alone or with either soluble laminin-1 or YIGSR 30 min before fixation. Forskolin (Sigma) was added to cultures at 20 μ M for the same period.

Cultures were fixed in acrolein (Sigma)²³, incubated in anti-cAMP antibody overnight at 4 °C, followed by 40 min incubation in Cy3-conjugated secondary antibody at 37 °C. cAMP immunofluorescence intensity was measured using IP Lab Spectrum P captured images under identical optics for all growth cones. The boundary of a phase-contrast image was superimposed on the fluorescent image of the same growth cone for measurement of growth-cone area. The fluorescence intensity of each growth cone was divided by its area to give a mean intensity for all growth cones in each culture, and mean values from 5 experiments were used. Statistical analysis (Student's *t*-test) compared cAMP intensity after addition of netrin-1 alone with controls (no additions) and with addition of netrin-1 with either laminin-1 or YIGSR.

Laminin-1/tubulin immunolocalization.

Cryostat sections of stage 37/38 retinas were incubated with anti-acetylated tubulin and a Cy3-conjugated secondary antibody followed by anti-laminin-1 and biotin-conjugated/streptavidin-fluorescein-conjugated secondary antibodies and viewed with a Leica confocal microscope.

Eye preparation.

The ectoderm and presumptive lens tissue were removed from the right eye primordia of *Xenopus* embryos at stage 27, ~1 h before axonogenesis begins. Unoperated left eyes served as internal controls. Peptides were added at 25–100 µg ml⁻¹ to Modified Barth's Solution. Embryos were fixed 6 h later at stage 32, and eyes were processed for wholemount immunohistochemistry using anti-acetylated tubulin. Images were captured with a CCD camera and defects in axon growth were scored on a scale of 0–3 (index of abnormality), with 0 being normal (Fig. 5c) and 3 representing severely disoriented growth around the ONH. Statistical analysis used Student's *t*-test.

Received 28 May; accepted 8 July 1999.

1. Bixby, J. L. & Harris, W. A. Molecular mechanisms of axon growth and guidance. *Annu. Rev. Cell Biol.* **7**, 117–159 (1991).
2. Culotti, J. G. & Kolodkin, A. L. Functions of netrins and semaphorins in axon guidance. *Curr. Opin. Neurobiol.* **6**, 81 (1996).
3. Tessier-Lavigne, M. & Goodman, C. S. The molecular biology of axon guidance. *Science* **274**, 1123–1133 (1996).
4. de la Torre, J. R. *et al.* Turning of retinal growth cones in a netrin-1 gradient mediated by the netrin receptor DCC. *Neuron* **19**, 1211–1224 (1997).
5. Deiner, M. S. *et al.* Netrin-1 and DCC mediate axon guidance locally at the optic disc: loss of function leads to optic nerve hypoplasia. *Neuron* **19**, 575–589 (1997).
6. Ming, G. L. *et al.* cAMP-dependent growth cone guidance by netrin-1. *Neuron* **19**, 1225–1235 (1997).
7. Cohen, J., Burne, J. F., Winter, J. & Bartlett, P. Retinal ganglion cells lose response to laminin with maturation. *Nature* **322**, 465–467 (1986).
8. Cohen, J., Burne, J. F., McKinlay, C. & Winter, J. The role of laminin and the laminin/fibronectin receptor complex in the outgrowth of retinal ganglion cell axons. *Dev. Biol.* **122**, 407–418 (1987).
9. Liesi, P. & Silver, J. Is astrocyte laminin involved in axon guidance in the mammalian CNS? *Dev. Biol.* **130**, 774–785 (1988).
10. de Curtis, I. & Reichardt, L. F. Function and spatial distribution in developing chick retina of the laminin receptor $\alpha\beta 1$ and its isoforms. *Development* **118**, 377–388 (1993).
11. Lilienbaum, A., Reszka, A. A., Horwitz, A. F. & Holt, C. E. Chimeric integrins expressed in retinal ganglion cells impair process outgrowth *in vivo*. *Mol. Cell. Neurosci.* **6**, 139–152 (1995).
12. Darribere, T., Yamada, K. M., Johnson, K. E. & Boucalt, J. C. The 140-kDa fibronectin receptor complex is required for mesodermal cell adhesion during gastrulation in the amphibian *Pleurodeles waltlii*. *Dev. Biol.* **126**, 182–194 (1988).
13. Powell, S. K. & Kleinman, H. K. Neuronal laminins and their cellular receptors. *Int. J. Biochem. Cell Biol.* **29**, 401–414 (1997).
14. Sephel, G. C. *et al.* Laminin A chain synthetic peptide which supports neurite outgrowth. *Biochem. Biophys. Res. Commun.* **162**, 821–829 (1989).
15. Rabacchi, S. A., Neve, R. L. & Drager, U. C. A positional marker for the dorsal embryonic retina is homologous to the high-affinity laminin receptor. *Development* **109**, 521–531 (1990).
16. Ardini, E. *et al.* Co-regulation and physical association of the 67-kDa monomeric laminin receptor and the $\alpha\beta 4$ integrin. *J. Biol. Chem.* **272**, 2342–2345 (1997).
17. Menard, S., Castronovo, V., Tagliabue, E. & Sobel, M. E. New insights into the metastasis-associated 67 kD laminin receptor. *J. Cell Biochem.* **67**, 155–165 (1997).
18. Weeks, B. S. *et al.* Adult and fetal human mesangial cells interact with specific laminin domains. *Am. J. Physiol.* **261**, F688–695 (1991).
19. Song, H. J., Ming, G. L. & Poo, M. M. cAMP-induced switching in turning direction of nerve growth cones. *Nature* **388**, 275–279 (1997); erratum, *ibid* **389**, 413 (1997).
20. Rothermel, J. D. & Parker Botelho, L. H. A mechanistic and kinetic analysis of the interactions of the diastereoisomers of adenosine 3',5'-(cyclic)phosphorothioate with purified cyclic AMP-dependent protein kinase. *Biochem. J.* **251**, 757–762 (1988).
21. Bates, C. A. & Meyer, R. L. Heterotrimeric G protein activation rapidly inhibits outgrowth of optic axons from adult and embryonic mouse, and goldfish retinal explants. *Brain Res.* **714**, 65–75 (1996).
22. Hauenberger, D., Klominek, J. & Sundqvist, K. G. Functional specialization of fibronectin-binding beta 1-integrins in T lymphocyte migration. *J. Immunol.* **153**, 960–971 (1994).
23. Wiemelt, A. P., Engleka, M. J., Skrupa, A. F. & McMorris, F. A. Immunohistochemical visualization and quantitation of cyclic AMP in single cells. *J. Biol. Chem.* **272**, 31489–31495 (1997).
24. Hanson, M. G. Jr, Shen, S., Wiemelt, A. P., McMorris, F. A. & Barres, B. A. Cyclic AMP elevation is sufficient to promote the survival of spinal motor neurons *in vitro*. *J. Neurosci.* **18**, 7361–7371 (1998).
25. Holt, C. E. A single-cell analysis of early retinal ganglion cell differentiation in *Xenopus*: from soma to axon tip. *J. Neurosci.* **9**, 3123–3145 (1989).
26. Lohof, A. M., Quillan, M., Dan, Y. & Poo, M. M. Asymmetric modulation of cytosolic cAMP activity induces growth cone turning. *J. Neurosci.* **12**, 1253–1261 (1992).
27. Song, H.-J. & Poo, M.-M. Signal transduction underlying growth cone guidance by diffusible factors. *Curr. Opin. Neurobiol.* **9**, 355–363 (1999).

28. Nieuwkoop, P. D. & Faber, J. *Normal Table of Xenopus laevis (Daudin). A Systematical and Chronological Survey of the Development from the Fertilized Egg till the End of Metamorphosis* (North-Holland, Amsterdam, 1967).
29. Harris, W. A., Holt, C. E., Smith, T. A. & Gallenson, N. Growth cones of developing retinal cells *in vivo*, on culture surfaces, and in collagen matrices. *J. Neurosci. Res.* **13**, 101–122 (1985).

Acknowledgements

We thank H. K. Kleinman, A. Wiemelt and J. Fawcett for peptides and antibodies, and B. Berninger, W. Harris, T. Gomez, S.-i. Ohnuma, B. Barres, S.-i. Nakagawa and T. Das for discussions. This work was supported by NIH, PEW Scholars Award and a MRC programme grant (C.H.) and NIH grants (M.-M.P., M.T.-L.).

Correspondence and requests for materials should be addressed to C.E.H. (e-mail: holt@mole.bio.cam.ac.uk).

Leptin reverses insulin resistance and diabetes mellitus in mice with congenital lipodystrophy

Ichiro Shimomura*, Robert E. Hammer†, Shinji Ikemoto*, Michael S. Brown* & Joseph L. Goldstein*

* Department of Molecular Genetics, † Department of Biochemistry and Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, Texas 75235-9046, USA

Congenital generalized lipodystrophy (CGL) is a rare autosomal recessive disorder characterized by a paucity of adipose (fat) tissue which is evident at birth and is accompanied by a severe resistance to insulin, leading to hyperinsulinaemia, hyperglycaemia and enlarged fatty liver¹. We have developed a mouse model that mimics these features of CGL²: the syndrome occurs in transgenic mice expressing a truncated version of a nuclear protein known as nSREBP-1c (for sterol-regulatory-element-binding protein-1c) under the control of the adipose-specific *aP2* enhancer. Adipose tissue from these mice was markedly deficient in messenger RNAs encoding several fat-specific proteins, including leptin², a fat-derived hormone that regulates food intake and energy metabolism³. Here we show that insulin resistance in our lipodystrophic mice can be overcome by a continuous systemic infusion of low doses of recombinant leptin, an effect that is not mimicked by chronic food restriction. Our results support the idea that leptin modulates insulin sensitivity and glucose disposal independently of its effect on food intake, and that leptin deficiency accounts for the insulin resistance found in CGL.

SREBP-1c, also known as ADD-1 (ref. 4), is one of three SREBPs that control transcription of genes encoding enzymes of lipid biosynthesis in animal cells⁵. The SREBPs are bound to membranes of the nuclear envelope and endoplasmic reticulum. In cholesterol-depleted cells, the transcriptionally active amino-terminal fragments of the SREBPs are released by proteolysis. They enter the nucleus and activate genes encoding enzymes involved in the synthesis of cholesterol and unsaturated fatty acids and in their uptake from low-density lipoprotein (LDL) particles. When sterols overaccumulate in cells, proteolysis is blocked and transcription of the target genes declines⁵.

The lipodystrophic transgenic mice express a truncated nuclear version of SREBP-1c, designated nSREBP-1c, which lacks a membrane attachment domain and so enters the nucleus in an unregulated way, without the need for proteolysis². Expression of nSREBP-1c is driven by the *aP2* promoter, which is specific for white and brown adipose tissue⁶. Although nSREBP-1c has been reported to

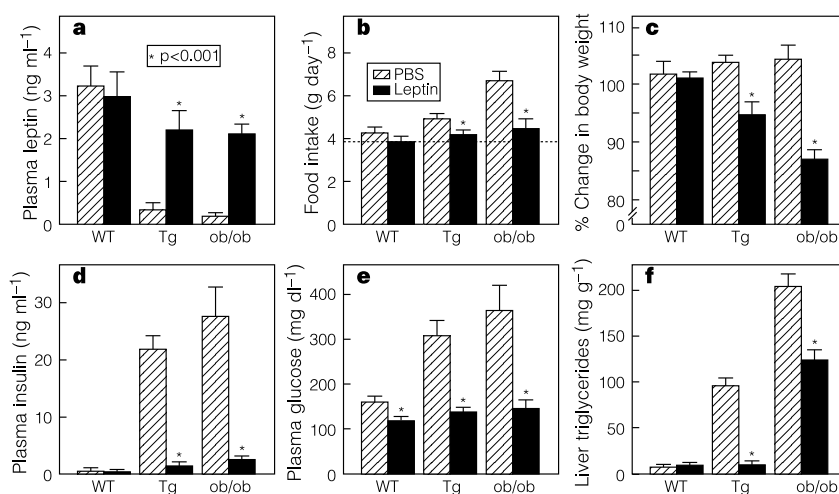


Figure 1 Metabolic parameters in wild-type (WT), transgenic aP2-nSREBP-1c436 (Tg), and *ob/ob* mice after continuous subcutaneous infusion of recombinant leptin (5 μ g per day for 12 days). On day 12, the mice were killed and the blood and liver removed. Each value denotes mean $\pm$ s.e.m. for groups of 4 mice. **a, d, e**, Plasma concentrations. **b**, Mean daily food intake averaged over 12 days after insertion of the micro-osmotic pump. Dashed line, amount of food per day given to the food-restricted group in Fig. 3.

c, Change in body weight after insertion of the pump. Mean '100% values' before insertion of the pump for wild-type, transgenic and *ob/ob* mice were 22.1, 25.9 and 50.1 g, respectively. **f**, Liver triglyceride content. Liver weights of control PBS-treated mice were 1.19 ± 0.06 , 2.22 ± 0.18 and 3.42 ± 0.36 g in wild-type, transgenic and *ob/ob* mice, respectively. In leptin-treated mice, the corresponding values were 1.01 ± 0.09 , 1.42 ± 0.10 and 2.29 ± 0.15 g, respectively. Asterisk, $P < 0.001$.

enhance adipocyte differentiation *in vitro*⁷, in our transgenic mice high levels of nSREBP-1c had the opposite effect². Adipocytes failed to differentiate fully, and the weights of white adipose tissue depots were markedly reduced. There was a severe reduction in mRNAs encoding proteins normally found in mature adipocytes, including the transcription factors C/EBP- α , and PPAR- γ and the adipocyte-specific proteins adipin and leptin. In contrast, there was an increase in the mRNA encoding Pref-1, a marker for undifferentiated adipocytes. The interscapular fat pad was enlarged, and its normal brown fat was replaced by immature-looking white fat. The mRNA for UCP-1, a marker for brown fat, was markedly reduced. The livers were massively enlarged and filled with triglyceride; other tissues, including the spleen, pancreas and lymph nodes, were also enlarged². Blood sugar was raised, despite plasma insulin being 60-fold above normal, and did not fall when large amounts of insulin were injected intraperitoneally. Insulin resistance did not seem to result from defective uptake of glucose in skeletal muscle: excised skeletal muscle from these animals showed a normal increase in glucose uptake when incubated with insulin *in vitro*².

A striking feature of adipose tissue in our aP2-nSREBP-1c transgenic mice was the 90% reduction in the amount of mRNA encoding leptin². A large reduction in plasma leptin also occurs in humans with CGL⁸. Leptin is secreted by adipocytes and acts on the hypothalamus to inhibit food intake and increase energy metabolism³; peripherally, it enhances fatty-acid oxidation⁹. A primary genetic deficiency of leptin occurs in *ob/ob* mice, which manifests as excessive food consumption, obesity, liver enlargement, insulin resistance and other metabolic abnormalities¹⁰. Treatment of *ob/ob* mice with leptin corrects the insulin resistance even before the animals have lost significant amounts of weight^{3,11}.

We investigated whether leptin deficiency could be the cause of the insulin resistance in our aP2-nSREBP-1c transgenic mice. We used an osmotic minipump to infuse leptin continuously into wild-type and *ob/ob* mice, a method previously used to demonstrate that a dose of 5 μ g per day of recombinant leptin can reverse the abnormalities in *ob/ob* mice but has no significant effect on wild-type animals¹¹.

Figure 1 presents metabolic parameters in wild-type, transgenic aP2-nSREBP-1c and *ob/ob* mice that were infused with phosphate-buffered saline (PBS) or recombinant leptin (5 μ g per day) for 12 days. aP2-nSREBP-1c mice and the *ob/ob* mice had very low plasma

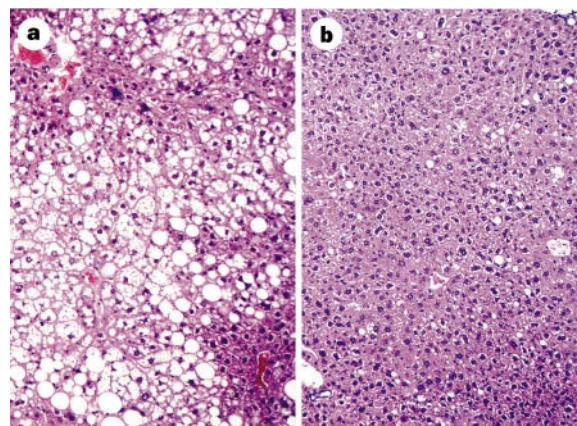


Figure 2 Representative histological sections of liver (haematoxylin and eosin stain) from transgenic aP2-nSREBP-1c436 mice. **a, b**, Mice were treated for 12 days with either PBS (**a**) or leptin (**b**). Original magnification, 200 $\times$.

leptin which was restored almost to normal by the infusion (Fig. 1a); this amount of exogenous leptin did not raise the leptin level in the plasma of wild-type mice. The aP2-nSREBP-1c mice ate $\sim 17\%$ more food than wild-type mice, and the *ob/ob* mice ate $\sim 60\%$ more (Fig. 1b). Leptin reduced food intake by 16 and 35% in aP2-nSREBP-1c and *ob/ob* mice, respectively. Body weights of the aP2-nSREBP-1c mice were about 10% higher than wild-type, despite the reduction in adipose tissue. The weight of *ob/ob* mice was more than twice that of wild-type mice (Fig. 1 legend). Leptin treatment reduced the weight of transgenic and *ob/ob* mice by 5 and 13%, respectively (Fig. 1c). Plasma insulin and glucose were markedly raised in aP2-nSREBP-1c and *ob/ob* mice, but both became almost normal after leptin treatment (Fig. 1d, e). Hepatic triglyceride content, which was abnormally high in aP2-nSREBP-1c mice, was corrected by leptin treatment (Fig. 1f). Hepatic triglycerides in *ob/ob* mice were also reduced substantially, but not to normal levels, by leptin treatment (Fig. 1f). These leptin-induced changes in hepatic triglyceride content were paralleled by changes in liver mass (Fig. 1 legend). The livers of aP2-nSREBP-1c mice showed massive vacuolization due to fat deposition (Fig. 2a); this was reversed by

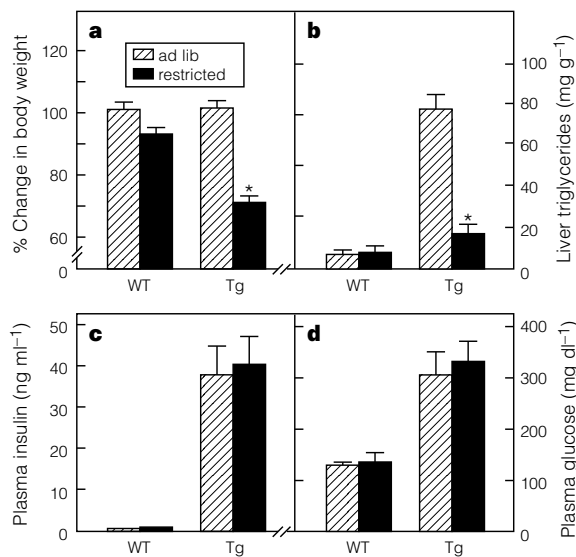


Figure 3 Metabolic parameters in wild-type (WT) and transgenic aP2-nSREBP-1c436 (Tg) mice after 12 days of food restriction. Baseline values were obtained by measuring food intake and body weight of 8 wild-type and 8 transgenic littermate mice every other day for 8 weeks. Daily food intake was 4.6 ± 0.1 and 5.4 ± 0.2 g in wild-type and transgenic mice, respectively (mean $\pm$ s.e.m.). Wild-type and transgenic mice were then divided into two weight-matched subgroups receiving *ad libitum* access to chow (shaded bars) or 3.8 g chow per day (black bars) (dashed line in Fig. 1b). All mice were killed 12 days after the start of food restriction. Each value denotes mean $\pm$ s.e.m. for 4 mice. **a**, Change in body weight after food restriction. Initial body weight was set at 100%. Mean '100% values' for wild-type and transgenic mice before food restriction were 23.9 and 30.2 g, respectively. **b**, Liver triglyceride content. **c**, **d**, Plasma concentrations. Asterisk, $P < 0.001$.

leptin treatment (Fig. 2b).

As leptin reduced food intake by the transgenic mice, it was possible that their metabolic improvement was due simply to this reduction in food intake. To rule this out, we severely restricted the food intake of a group of aP2-nSREBP-1c transgenic mice for 12 days (Fig. 3), having recorded the intake by wild-type and transgenic animals beforehand (average, 4.6 and 5.4 g per day). We reduced the intake by both groups to 3.8 g per day, which represents a 30% reduction for aP2-nSREBP-1c mice and is a greater reduction than the 16% cut in food intake seen with leptin treatment (Fig. 1b). The body weight of food-restricted aP2-nSREBP-1c mice declined by 28%, which is consistent with the severe caloric restriction (Fig. 3a); liver triglyceride content also fell substantially (Fig. 3b). Despite these signs of caloric deprivation, neither plasma insulin nor glucose fell in the aP2-nSREBP-1c transgenic mice (Fig. 3c, d).

Leptin did not act by reversing the lipodystrophy in aP2-nSREBP-1c transgenic mice: omental fat remained visibly low after leptin treatment, and the marker adipose-specific mRNAs for PPAR- γ , adipon and UCP-1 were still barely detectable in white and brown fat (Fig. 4a). Moreover, a raised level of TNF- α mRNA in the white and brown fat of transgenic lipodystrophic mice² was not normalized by leptin treatment (Fig. 4b).

Our results indicate that the insulin resistance in this mouse model of lipodystrophy is caused by a deficiency of leptin that is secondary to a failure of adipocyte differentiation. Insulin sensitivity can be restored by subcutaneous infusion of physiological amounts of leptin. Although the mechanism by which leptin deficiency causes insulin resistance in these mice is unknown, it is not due to ingestion of excess calories because caloric restriction substantially reduced body weight without reversing the hyperinsulinaemia or hyperglycaemia.

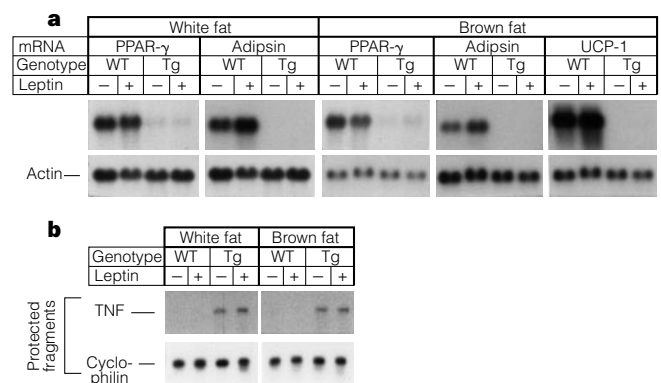


Figure 4 Amounts of mRNAs in white fat and brown fat of wild-type (WT) and transgenic aP2-nSREBP-1c (Tg) mice treated with leptin. Total RNA was isolated from adipose tissue of the mice used for Fig. 1 and samples from 4 mice were pooled. **a**, Northern-blot analysis of adipogenic mRNAs. Aliquots of pooled RNA (10 μ g) were subjected to electrophoresis and blot hybridization with the indicated ³²P-labelled cDNA probe. Blots were exposed to Reflection⁴⁹⁶ film (DuPont-NEN) at -80 °C for 16 h (adipon and UCP) or 24 h (PPAR- γ). **b**, RNase protection assay. Aliquots of pooled RNA (10 μ g) were hybridized in solution for 10 min at 68 °C to a ³²P-labelled cRNA probe for TNF- α in the presence of a cRNA probe for cyclophilin. After RNase digestion, the protected fragments were separated by electrophoresis and exposed to film at -80 °C for 48 h (TNF- α) or 2 h (cyclophilin).

It has been suggested that one function of leptin could be to divert fatty acids into adipose tissue for storage, and away from other tissues, including the insulin-secreting β -cells of the pancreatic islets of Langerhans⁹, which in the absence of leptin might take up more fatty acids for storage or metabolism and become resistant to the stimulatory effects of insulin on glucose uptake and metabolism¹². The diversion of fatty acids to peripheral tissues might be even worse in aP2-nSREBP-1c transgenic mice than in *ob/ob* mice because of their lack of differentiated adipocytes. It remains to determine the triglyceride content of pancreatic β -cells and other cells in the lipodystrophic mice before and after leptin treatment, and whether leptin in this model is working solely on the hypothalamus or also has a direct action on peripheral tissues.

In another transgenic mouse model of lipodystrophy¹³, created by overexpression of a dominant-negative form of C/EBP α , a transcription factor necessary for normal adipocyte differentiation, the mice were also insulin-resistant, and were hyperinsulinaemic, hyperglycaemic and had fatty livers; the amount of free, active leptin in plasma was reduced to ~5% of that in wild-type mice. It will be interesting to see whether the insulin resistance of these mice also improves in response to leptin infusion.

Humans with CGL have extremely low levels of plasma leptin⁸, so a clinical trial of leptin treatment in these patients might be warranted. □

Methods

Animals.

Transgenic mice expressing SREBP-1c436 (encoding amino acids 1–436 of human SREBP-1c) in adipose tissue under the control of the adipocyte-specific *aP2* promoter/enhancer have been described². The A line was used in these studies. Wild-type mice (C57BL/6J $\times$ SJL) were littermates of transgenic SREBP-1c436. 'Obese' (C57BL/6J-*ob/ob*) mice were purchased from The Jackson Laboratory (Bar Harbor, Maine). All aP2-nSREBP-1c transgenic and wild-type littermate mice used in this study were female; all *ob/ob* mice were male. Mice were housed in colony cages on a 14-h light (07:00–21:00 h)/10-h day (21:00–07:00 h) cycle and were maintained on Teklad 4% mouse/rat diet 7002 (Harlan Teklad Premier Laboratory Diets, Madison).

Metabolic assays and procedures.

Blood was drawn from the retro-orbital sinus and the plasma separated immediately and stored at -70 °C. Plasma glucose was measured by using a glucose (Trinder)-100 kit

(Sigma Diagnostics), plasma insulin by a monoclonal anti-rat insulin radioimmunoassay with a rat insulin RIA kit (Linco), and plasma leptin by a monoclonal anti-mouse leptin radioimmunoassay with a mouse leptin RIA kit (Linco). Liver triglyceride was measured as described¹⁴. For histology, livers were fixed in 10% (v/v) buffered formalin, dehydrated through graded ethanol concentrations and embedded in paraffin; sections were cut 4 μ m thick and stained with haematoxylin and eosin.

Leptin treatment.

All mice were studied at 14–16 weeks of age and were housed individually with continuous free access to chow and water. Body weight and food intake were recorded every other day at 08:00 h. After 3 weeks of baseline measurements of body weight and food intake, each group of mice was divided into two weight-matched subgroups, and an Alzet micro-osmotic pump (model 1002, Alza) was inserted subcutaneously in the back of each mouse¹¹. The pumps delivered 0 or 5 μ g of mouse recombinant leptin (Linco) per day in a total volume of 0.1 ml of Dulbecco's PBS (Gibco BRL). After pump placement, measurements of food intake and body weight were continued. All mice were killed and blood was drawn early in the light cycle 12 days after inserting the pumps. Experiments were repeated once, with similar results.

Food restriction.

Wild-type and transgenic aP2–nSREBP-1c436 littermate mice were studied at 22–24 weeks old. All mice were housed individually with continuous access to chow and water. Body weights and food intakes were recorded every other day at 08:00 h. After 8 weeks of baseline measurements of body weight and food intake, each group of mice was divided into two weight-matched subgroups which were given either *ad libitum* access to chow or 3.8 g chow per day. The 3.8-g chow diet was given once daily at 08:00 h. All mice were killed and blood was drawn early in the light cycle 12 days after starting food restriction. Experiments were repeated once, with virtually identical results.

Adipose-tissue mRNA.

Mice were killed and adipose tissue was obtained from the parametrium (white fat) and the interscapular fat pad (brown fat). Total RNA was isolated and aliquots analysed by northern blotting or RNase protection assay with ³²P-labelled cDNA or cRNA progress as described^{2,15}.

Received 7 May; accepted 13 July 1999.

- Seip, M. & Trygstad, O. Generalized lipodystrophy, congenital and acquired (lipoatrophy). *Acta Paediatr. Suppl.* **413**, 2–28 (1996).
- Shimomura, I. *et al.* Insulin resistance and diabetes mellitus in transgenic mice expressing nuclear SREBP-1c in adipose tissue: model for congenital generalized lipodystrophy. *Genes Dev.* **12**, 3182–3194 (1998).
- Friedman, J. M. & Halaas, J. L. Leptin and the regulation of body weight in mammals. *Nature* **395**, 763–770 (1998).
- Tontonoz, P., Kim, J. B., Graves, R. A. & Spiegelman, B. M. ADD1: a novel helix–loop–helix transcription factor associated with adipocyte determination and differentiation. *Mol. Cell. Biol.* **13**, 4753–4759 (1993).
- Brown, M. S. & Goldstein, J. L. The SREBP pathway: Regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. *Cell* **89**, 331–340 (1997).
- Ross, S. R. *et al.* A fat-specific enhancer is the primary determinant of gene expression for adipocyte P2 *in vivo*. *Proc. Natl Acad. Sci. USA* **87**, 9590–9594 (1990).
- Kim, J. B. & Spiegelman, B. M. ADD1/SREBP1 promotes adipocyte differentiation and gene expression linked to fatty acid metabolism. *Genes Dev.* **10**, 1096–1107 (1996).
- Pardini, V. C. *et al.* Leptin levels, β -cell function, and insulin sensitivity in families with congenital and acquired generalized lipodystrophic diabetes. *J. Clin. Endocrinol. Metab.* **83**, 503–508 (1998).
- Unger, R. H., Zhou, Y.-T. & Orci, L. Regulation of fatty acid homeostasis in cells: Novel role of leptin. *Proc. Natl Acad. Sci. USA* **96**, 2327–2332 (1999).
- Zhang, Y. *et al.* Positional cloning of the mouse *obese* gene and its human homologue. *Nature* **372**, 425–432 (1994).
- Harris, R. B. S. *et al.* A leptin dose–response study in obese (*ob/ob*) and lean (*+/+*) mice. *Endocrinology* **139**, 8–19 (1998).
- McGarry, J. D. What if Minkowski had been ageusic? An alternative angle on diabetes. *Science* **258**, 766–770 (1992).
- Moitra, J. *et al.* Life without white fat: a transgenic mouse. *Genes Dev.* **12**, 3168–3181 (1998).
- Yokode, M., Hammer, R. E., Ishibashi, S., Brown, M. S. & Goldstein, J. L. Diet-induced hypercholesterolemia in mice: Prevention by overexpression of LDL receptors. *Science* **250**, 1273–1275 (1990).
- Shimomura, I., Shimano, H., Korn, B. S., Bashmakov, Y. & Horton, J. D. Nuclear sterol regulatory element binding proteins activate genes responsible for entire program of unsaturated fatty acid biosynthesis in transgenic mouse liver. *J. Biol. Chem.* **273**, 35299–35306 (1998).

Acknowledgements

This work was supported by grants from the NIH, the Moss Heart Foundation, and the Perot Family foundation. I.S. held a research fellowship from the Manpei Suzuki Diabetes Foundation of Tokyo; S. Ikemoto holds a research fellowship from the National Institute of Health and Nutrition of Japan. We thank J. D. McGarry and R. Unger for their suggestions; R. Gibson, B. Hinnant, A. Cox and M. Villarreal for help with the animal studies; S. Clark and R. Craddock for technical assistance; and members of J. A. Richardson's laboratory for histology.

Correspondence and requests for materials should be addressed to M.S.B.

(e-mail: mbrow1@mednet.swmed.edu) or J.L.G. (e-mail: jgold@mednet.swmed.edu).

Rapid on/off cycling of cytokine production by virus-specific CD8⁺ T cells

Mark K. Slifka, Fernando Rodriguez & J. Lindsay Whitton

Department of Neuropharmacology, CVN-9, The Scripps Research Institute, 10550 N. Torrey Pines Rd., La Jolla, California 92037, USA

CD8-positive T cells protect the body against viral pathogens by two important mechanisms: production of antiviral cytokines^{1,2} and lysis of infected cells^{3,4}. Cytokine production can have both local and systemic consequences^{5,6}, whereas cytolytic activity is limited to infected cells that are in direct contact with T cells^{7–9}. Here we analyse activated CD8-positive T cells from mice infected with lymphocytic choriomeningitis virus and find that cytokines are not produced *ex vivo* in the absence of peptide stimulation, but that they are rapidly generated after T cells encounter viral peptides bound to the major histocompatibility complex. Remarkably, cytokine production ceases immediately upon dissociation of the T cells from their targets and resumes when antigenic contact is restored. In contrast to the 'on/off/on' cycling of cytokines, the pore-forming cytotoxic protein perforin is constitutively maintained. Our results indicate that there is differential expression of effector molecules according to whether the antiviral product is secreted (like cytokines) or stored inside the cell (like perforin). The ability to turn cytokines on and off while maintaining intracellular stores of perforin shows the

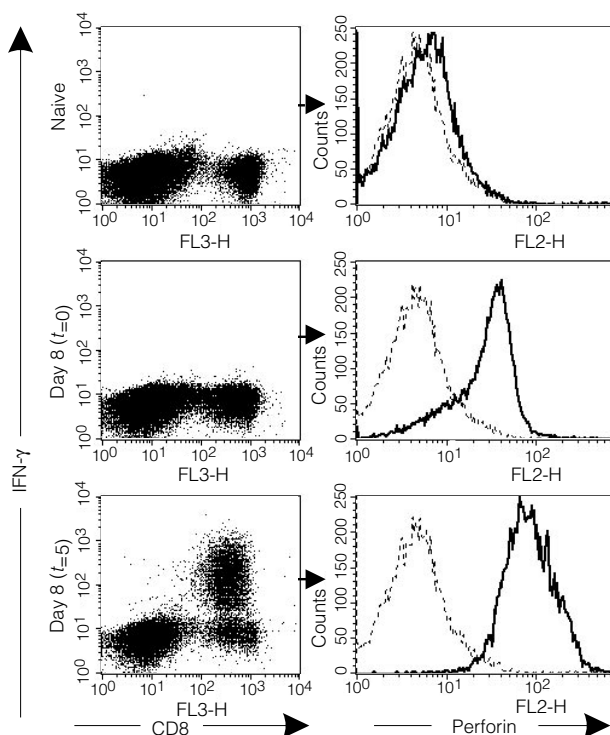


Figure 1 Perforin expression in virus-specific CD8⁺ T cells. Intracellular perforin levels were measured in spleen cells from naive mice or from LCMV-infected mice (8 days post-infection) directly *ex vivo* ($t = 0$) or after 5 h of stimulation with NP_{118–126} peptide. Spleen cells from perforin-deficient (PKO) mice were stained as a negative control (dashed lines). The perforin mean fluorescence intensity (MFI) of each gated cell population is shown in the top right corner of the histograms (PKO CD8⁺ T-cell MFI = 8.8). Note that directly *ex vivo* ($t = 0$), CD8⁺ T cells contain perforin but not IFN- γ , but after peptide stimulation, virus-specific T cells shift from perforin⁺IFN- γ [–] to perforin⁺IFN- γ ⁺ ($t = 5$ h).

versatility of the cellular immune response and provides a mechanism for maintaining effective immune surveillance while reducing systemic immunopathology.

CD8-positive T cells are required for the control and clearance of many viruses¹⁰, including lymphocytic choriomeningitis virus (LCMV)^{11–15}. LCMV infection results in a massive expansion of CD8⁺ T cells^{16,17} and, in BALB/c mice, ~50% of these cells are virus-specific by 8 days post-infection¹⁷. We used flow cytometry to compare perforin levels and cytokine expression in activated CD8⁺ T cells at the peak of the antiviral immune response (Fig. 1). Naive CD8⁺ T cells expressed little or no perforin, whereas about 80% of the CD8⁺ T cells from mice at 8 days post-infection expressed perforin directly *ex vivo* in the absence of any *in vitro* peptide stimulation (*t* = 0). We confirmed these findings by western blot analysis, and showed that these day-8 cells could lyse peptide-coated target cells in the presence of cycloheximide, indicating that *de novo* protein synthesis was not required (see Supplementary information). Stimulation of CD8⁺ T cells with the immunodominant peptide¹⁸ (LCMV NP_{118–126}) for 5 h *in vitro* induced ~60% of the CD8⁺ cells to produce high levels of interferon- γ (IFN- γ); by gating on CD8⁺IFN- γ ⁺ T cells, we found that >99% of these virus-specific T cells contained intracellular stores of perforin.

Unlike perforin, IFN- γ was undetectable in CD8⁺ T cells analysed directly *ex vivo* at 8 days post-infection (Fig. 1, *t* = 0), suggesting that these cells do not produce cytokines constitutively *in vivo*. To determine how rapidly CD8⁺ T cells could initiate cytokine production after antigen contact, spleen cells were stimulated *in vitro* with peptide. Virus-specific CD8⁺ T cells began to produce both IFN- γ and tumour necrosis factor- α (TNF- α) within 30 min after peptide stimulation, and the maximum number of cytokine-producing T

cells was reached within 6 h (Fig. 2a). IFN- γ production was completely abolished by the addition of actinomycin D (Act D), indicating that production of this cytokine required *de novo* RNA synthesis (Fig. 2b). In contrast, a small number of CD8⁺ T cells produced TNF- α in the presence of Act D, although the numbers were reduced by ~60% at 30 min post-stimulation and declined thereafter. This result indicated that pre-existing TNF- α RNA contributed to the initial burst of TNF- α production, but was not sufficient to drive the high levels of sustained TNF- α production observed during the first 6 h of stimulation. Transient TNF- α production from preformed RNA has also been observed following *in vitro* stimulation of CD4⁺ cells with phorbol myristate acetate (PMA) and ionomycin in the presence of Act D¹⁹. As expected, the addition of cycloheximide to the cultures completely suppressed both IFN- γ and TNF- α synthesis (Fig. 2b). Thus, during a brisk antiviral immune response, IFN- γ and TNF- α are not constitutively expressed by virus-specific CD8⁺ T cells; instead, cytokine production is initiated only when T cells come into direct contact with specific peptide antigen.

It is generally assumed that, upon full activation, virus-specific T-cell effector functions are constitutively expressed and remain so until the virus infection has been eradicated. However, our results (Figs 1, 2) and those of other studies^{17,20} have shown that, in the absence of stimulation, CD8⁺ T cells are not expressing cytokines when analysed directly *ex vivo*. The significance of this finding has not been previously appreciated, nor has the underlying mechanism been described. Our identification of highly cytolytic CD8⁺ T cells that were not actively producing cytokines led us to propose that virus-specific T cells could downregulate cytokine production without losing cytolytic functions. Therefore, we examined the effects of disrupting T-cell–antigen-presenting cell (APC) contact in CD8⁺ T cells that had been stimulated with peptide-coated APCs

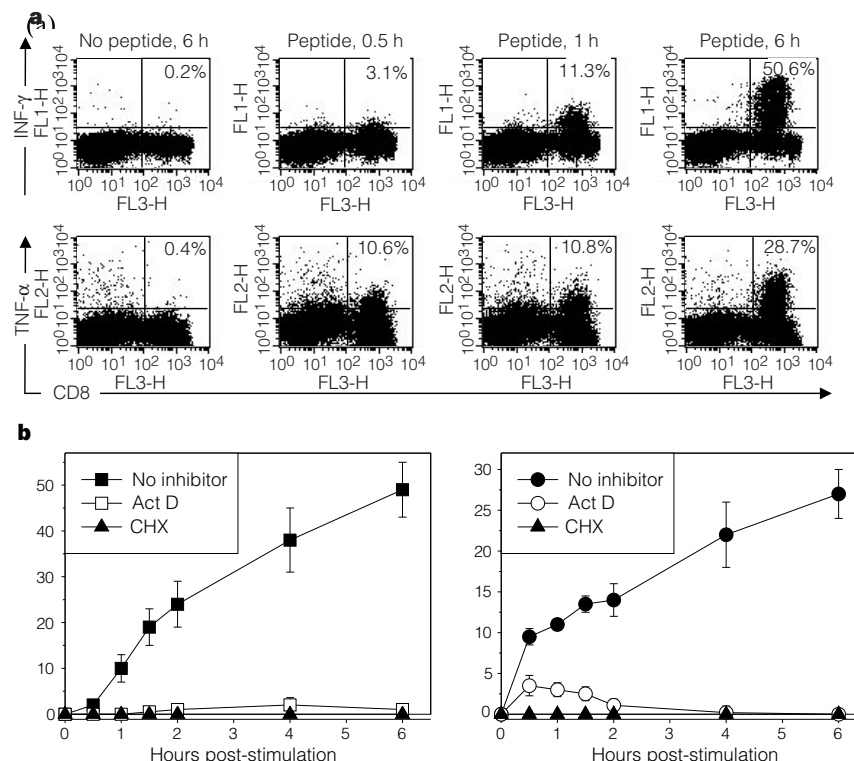


Figure 2 Kinetics of virus-specific cytokine production. **a**, Spleen cells from mice at 8 days post-infection were cultured directly *ex vivo* in the presence or absence of LCMV-NP_{118–126}, and both IFN- γ and TNF- α levels were quantified by intracellular cytokine staining. Numbers in the top right quadrants indicate the percentages of CD8⁺ T cells expressing IFN- γ or TNF- α . Minimal cytokine production was observed in the absence of

peptide stimulation (No peptide, 6 h). **b**, To determine whether pre-existing cytokine mRNA was involved in the rapid kinetics of IFN- γ and TNF- α production, spleen cells were stimulated with peptide in the presence or absence of Act D. Cytokine production was not observed after peptide stimulation of naive T cells (data not shown and ref. 17) indicating that only virus-specific T-cell responses are detected by this assay. CHX, cycloheximide.

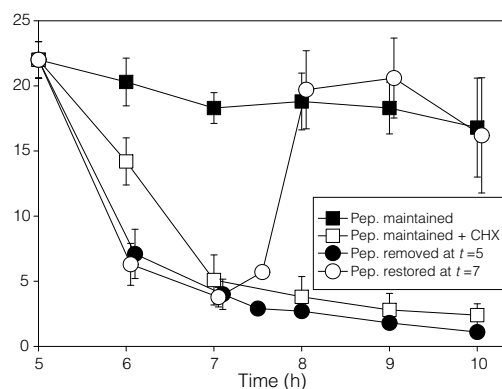


Figure 3 Rapid on/off/on cycling of cytokine production by activated CD8⁺ T cells. Spleen cells from LCMV-infected mice were stimulated for 5 h with peptide-coated biotinylated BALB clone 7 cells (Pep.). At this time point ($t = 5$) CD8⁺ T cells produced high levels of IFN- γ and were either unmanipulated (filled squares), treated with cycloheximide (open squares), or were separated from APC contact (filled and open circles). The recovered effector cells were immediately transferred to wells containing APCs in the absence of peptide. Within two hours, cytokine production dropped to baseline levels (filled and open circles; $t = 7$ h). One group (open circles) was restimulated with peptide (10^{-7} M) to determine how rapidly T cells could re-initiate cytokine production. The data show the average and standard deviation of three mice assayed individually.

for 5 h to maximize active cytokine production (Fig. 3). If the cultures were not manipulated further, IFN- γ production was maintained at steady-state levels (filled squares). Addition of cycloheximide to the cultures at $t = 5$ h (open squares) resulted in abrogation of protein synthesis, and the number of IFN- γ ⁺ T cells dropped to baseline levels within 2 h. If the T-cell-APC was disrupted and peptide-coated APCs were removed (filled circles), cytokine production declined at a rate similar to that observed following the addition of the protein-synthesis inhibitor cycloheximide. This result indicates that IFN- γ production is terminated immediately upon disruption of contact between the CD8⁺ T cell and its cognate antigen, and dovetails well with the notorious instability of cytokine mRNAs²¹, whose rapid degradation may facilitate cytokine down-regulation. To determine how rapidly T cells could re-initiate IFN- γ production, free peptide (10^{-7} M) was added to one group of APC-depleted cultures (open circles) after the number of cytokine-producing CD8⁺ T cells had fallen to baseline levels ($t = 7$ h). Within 60 min, the number of IFN- γ ⁺ T cells rebounded to the original levels observed before APC removal. This is in contrast to the 4–5 h of stimulation required to maximize the IFN- γ response of T cells analysed directly *ex vivo* (Fig. 2).

This is, to our knowledge, the first study to show that activated CD8⁺ T cells can cycle cytokine production on, off and on again in a regulated, antigen-specific manner while maintaining constitutive cytolytic activity. What might be the biological significance of this observation? Cytokines are responsible for many of the symptoms of virus infection, and cytokine release can be harmful and sometimes lethal to the host^{5,22–26}. As cytokines are rapidly secreted, constitutive expression of these proteins would lead to their inappropriate release in non-infected tissues as activated T cells migrated through various organs or the bloodstream. By initiating cytokine production only upon antigen contact, CD8⁺ T cells minimize the immunopathology that would occur if these molecules were secreted in a constitutive manner. Furthermore, by limiting cytokine production to periods of direct contact with virally infected targets, CD8⁺ T cells focus their resources at the site of infection. By increasing our understanding of these and other mechanisms that govern T-cell functions during acute and chronic infections, we may be able to develop better therapeutic strategies that will reduce potentially autoreactive immune responses and alleviate many T-cell-mediated symptoms of disease. □

Methods

Intracellular perforin staining.

Adult BALB/c mice were injected intraperitoneally with 2×10^5 plaque-forming units of LCMV-Armstrong. Directly *ex vivo*, cells were either temporarily stored at 4 °C or were incubated for 5 h at 37 °C, 6% CO₂ with 10^{-7} M NP_{118–126} peptide in RPMI containing 10% FBS, L-glutamine and $2 \mu\text{g ml}^{-1}$ brefeldin A (Sigma). To detect intracellular perforin (Fig. 1), cells were fixed with 2% formaldehyde in PBS for 20 min on ice, washed and permeabilized with 0.1% saponin (Sigma) in PBS containing 1% FBS. Anti-perforin antibody (clone P1-8, Kamiya Biomedical) was diluted 1:200 in 0.3% saponin/PBS containing 5% normal goat serum, and added to the cells. Following a 30 min incubation on ice, the cells were washed with 0.1% saponin and incubated with polyclonal goat anti-rat Ig-PE for 30 min. After washing, cells were stained for 30 min with anti-CD8-cychrome and anti-IFN- γ -FITC (Pharmingen). The cells were washed with 0.1% saponin, then with PBS + 5% FBS, and were stored at 4 °C in PBS containing 2% formaldehyde until analysis. Cells were acquired on a FACScan flow cytometer (80,000–100,000 events per sample) and analysed using CELLQUEST software (Becton Dickinson). The data shown are representative of three experiments.

Intracellular cytokine staining.

We evaluated the kinetics of cytokine production (Fig. 2a) by stimulating d8 cells with peptide for the indicated times. Brefeldin A was added at a final concentration of $2 \mu\text{g ml}^{-1}$ for the entire culture period (0.5–6 h). To evaluate the importance of mRNA synthesis in the cytokine response (Fig. 2b), spleen cells were pre-warmed at 37 °C for 20 min in the presence or absence of Act D ($50 \mu\text{g ml}^{-1}$) before the addition of peptide. In parallel, spleen cells were pre-warmed in the presence of cycloheximide (CHX; $100 \mu\text{g ml}^{-1}$) for 20 min before the addition of peptide. The addition of this protein synthesis inhibitor completely blocked cytokine production. For staining, cells were immediately placed on ice, washed and stained for CD8 (Pharmingen) according to the manufacturer's directions. FITC-conjugated anti-IFN- γ and PE-conjugated anti-TNF- α monoclonal antibodies were purchased from Pharmingen. The data show the average and standard deviation of three mice assayed individually.

APC depletion.

For analysis of the termination and re-initiation of cytokine production (Fig. 3), we stimulated d8 spleen cells (10^6 per well) with peptide-coated biotinylated BALB cl7 cells (2×10^5 per well). Brefeldin A was added to a final concentration of $2 \mu\text{g ml}^{-1}$ for the final 60 min of culture to facilitate intracellular cytokine accumulation. T-cell-APC contact was disrupted at the indicated time points by vigorous pipetting, and we removed the peptide-coated stimulator cells using a novel magnetic-bead depletion procedure. To prepare APCs, BALB clone 7 cells were washed once with PBS, resuspended at 10^7 per ml and biotinylated in PBS supplemented with 1 mM MgCl₂, 0.1 mM CaCl₂ and 1.0 mg ml^{-1} freshly added biotin (Calbiochem) for 20 min on ice. Biotinylation was quenched by adding RPMI containing 10% FBS and cells were washed twice before resuspension in medium containing 10^{-7} M peptide (LCMV NP_{118–126}). Cells were incubated at 37 °C for 1 h and washed extensively to remove unbound peptide. Biotinylation of targets had no effect on peptide presentation over a range of peptide doses (10^{-5} M to 10^{-8} M). Spleen cells (10^6 per well) from LCMV infected mice were incubated (37 °C, 6% CO₂) with 2×10^5 biotinylated fibroblasts coated with 10^{-7} M peptide. After 5 h of incubation to activate the maximal number of cytokine-producing cells, streptavidin-coated magnetic beads (Perceptive Biosystems) were added and mixed by pipetting. After 20 min at 37 °C, the cultures were resuspended by vigorous pipetting, and biotinylated cells were removed with a magnet (Perceptive Biosystems). More than 95% of the target cells were removed by this depletion procedure. About 4×10^5 of the recovered CD8⁺ T cells were transferred to wells containing 4×10^5 uncoated APCs and analysed for IFN- γ production as described in the text. The immediate cessation of IFN- γ synthesis by CD8⁺ T cells that occurs after APC removal was not a consequence of mechanical damage incurred during APC depletion, because the depletion procedure had no effect on IFN- γ production that had been stimulated by incubation with antibodies to CD3 and CD28 (data not shown). At each time point, cells were removed from the culture plates, washed, and for convenience placed at 4 °C overnight in PBS containing 5% FBS and anti-CD8 (Pharmingen). We detected IFN- γ production as described above.

Received 13 June; accepted 5 July 1999.

- Ramsay, A. J., Ruby, J. & Ramshaw, I. A. A case for cytokines as effector molecules in the resolution of virus infection. *Immunol. Today* **14**, 155–157 (1993).
- Guidotti, L. G. & Chisari, F. V. To kill or to cure: options in host defense against viral infection. *Curr. Opin. Immunol.* **8**, 478–483 (1996).
- Walsh, C. M. et al. Immune function in mice lacking the perforin gene. *Proc. Natl Acad. Sci. USA* **91**, 10854–10858 (1994).
- Kagi, D. et al. Cytotoxicity mediated by T cells and natural killer cells is greatly impaired in perforin-deficient mice. *Nature* **369**, 31–37 (1994).
- Vassalli, P. The pathophysiology of tumor necrosis factors. *Annu. Rev. Immunol.* **10**, 411–452 (1992).
- Tracey, K. J. et al. Shock and tissue injury induced by recombinant human cachectin. *Science* **234**, 470–474 (1986).
- Lowin, B., Hahne, M., Mattmann, C. & Tschopp, J. Cytolytic T-cell cytotoxicity is mediated through perforin and Fas lytic pathways. *Nature* **370**, 650–652 (1994).
- Bossi, G. & Griffiths, G. M. Degranulation plays an essential part in regulating cell surface expression of Fas ligand in T cells and natural killer cells. *Nature Med.* **5**, 90–96 (1999).
- Hahn, K. et al. Antigen presentation and cytotoxic T lymphocyte killing studied in individual, living cells. *Virology* **201**, 330–340 (1994).
- Whitton, J. L. & Oldstone, M. B. A. in *Fields' Virology* Vol. 3 (eds Fields, B. N., Knipe, D. M. & Howley,

- P. M.) 345–374 (Raven, New York, 1996).
- Ahmed, R. *et al.* Selection of genetic variants of lymphocytic choriomeningitis virus in spleens of persistently infected mice: role in suppression of cytotoxic T lymphocyte response and viral persistence. *J. Exp. Med.* **160**, 521–540 (1984).
 - Byrne, J. A. & Oldstone, M. B. A. Biology of cloned cytotoxic T lymphocytes specific for lymphocytic choriomeningitis virus: clearance of virus in vivo. *J. Virol.* **51**, 682–686 (1984).
 - Moskophidis, D., Cobbold, S. P., Waldmann, H. & Lehmann-Grube, F. Mechanism of recovery from acute virus infection: treatment of lymphocytic choriomeningitis virus-infected mice with monoclonal antibodies reveals that *Lyt-2⁺* T lymphocytes mediate clearance of virus and regulate the antiviral antibody response. *J. Virol.* **61**, 1867–1874 (1987).
 - Whitton, J. L. Lymphocytic choriomeningitis virus CTL. *Semin. Virol.* **1**, 257–262 (1990).
 - Schulz, M. *et al.* Major histocompatibility complex-dependent T cell epitopes of lymphocytic choriomeningitis virus nucleoprotein and their protective capacity against viral disease. *Eur. J. Immunol.* **19**, 1657–1668 (1989).
 - Butz, E. A. & Bevan, M. J. Massive expansion of antigen-specific CD8⁺ T cells during an acute virus infection. *Immunity* **8**, 167–175 (1998).
 - Murali-Krishna, K. *et al.* Counting antigen-specific CD8 T cells: a reevaluation of bystander activation during viral infection. *Immunity* **8**, 177–187 (1998).
 - Whitton, J. L. *et al.* Molecular analyses of a five-amino-acid cytotoxic T-lymphocyte (CTL) epitope: an immunodominant region which induces nonreciprocal CTL cross-reactivity. *J. Virol.* **63**, 4303–4310 (1989).
 - Yang, Y., Chang, J. F., Parnes, J. R. & Fathman, C. G. T cell receptor (TCR) engagement leads to activation-induced splicing of tumor necrosis factor (TNF) nuclear pre-mRNA. *J. Exp. Med.* **188**, 247–254 (1998).
 - Flynn, K. J. *et al.* Virus-specific CD8⁺ T cells in primary and secondary influenza pneumonia. *Immunity* **8**, 683–691 (1998).
 - Sachs, A. B. Messenger RNA degradation in eukaryotes. *Cell* **74**, 413–421 (1993).
 - Hayden, F. G. *et al.* Local and systemic cytokine responses during experimental human influenza A virus infection. Relation to symptom formation and host defense. *J. Clin. Invest.* **101**, 643–649 (1998).
 - Campbell, I. L. Neuropathogenic actions of cytokines assessed in transgenic mice. *Int. J. Dev. Neurosci.* **13**, 275–284 (1995).
 - Masuda, N. *et al.* Interleukin 6 is a cause of flu-like symptoms in treatment with a deoxycytidine analogue. *Br. J. Cancer* **78**, 388–389 (1998).
 - Selin, L. K., Varga, S. M., Wong, I. C. & Welsh, R. M. Protective heterologous antiviral immunity and enhanced immunopathogenesis mediated by memory T cell populations. *J. Exp. Med.* **188**, 1705–1715 (1998).
 - Tracey, K. J. & Cerami, A. Tumor necrosis factor and regulation of metabolism in infection: role of systemic versus tissue levels. *Proc. Soc. Exp. Biol. Med.* **200**, 233–239 (1992).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank A. Lord for administrative assistance. The work was supported by grants from the NIH (J.L.W. and M.K.S.) and by Eusko Juarlaritza (F.R.).

Correspondence and requests for materials should be addressed to M.K.S. (e-mail: mslifka@scripps.edu).

Cell transformation by the superoxide-generating oxidase Mox1

Young-Ah Suh^{*,‡}, Rebecca S. Arnold^{*,‡}, Bernard Lassegue[†], Jing Shi^{*,}
Xiangxi Xu^{*,} Dan Sorescu[†], Andrew B. Chung[†], Kathy K. Griendling[†]
& J. David Lambeth^{*}

Departments of ^{*}Biochemistry and [†]Medicine, Emory University Medical School, Atlanta, Georgia 30322, USA

[‡]These authors contributed equally to this work.

Reactive oxygen species (ROS) generated in some non-phagocytic cells are implicated in mitogenic signalling and cancer^{1–6}. Many cancer cells show increased production of ROS⁷, and normal cells exposed to hydrogen peroxide or superoxide show increased proliferation⁸ and express growth-related genes^{9–11}. ROS are generated in response to growth factors, and may affect cell growth^{2,3,12,13}, for example in vascular smooth-muscle cells^{6,13–15}. Increased ROS in *Ras*-transformed fibroblasts correlates with increased mitogenic rate¹⁶. Here we describe the cloning of *mox1*, which encodes a homologue of the catalytic subunit of the superoxide-generating NADPH oxidase of phagocytes^{17,18}, gp91phox. *mox1* messenger RNA is expressed in colon, prostate,

uterus and vascular smooth muscle, but not in peripheral blood leukocytes. In smooth-muscle cells, platelet-derived growth factor induces *mox1* mRNA production, while antisense *mox1* mRNA decreases superoxide generation and serum-stimulated growth. Overexpression of *mox1* in NIH3T3 cells increases superoxide generation and cell growth. Cells expressing *mox1* have a transformed appearance, show anchorage-independent growth and produce tumours in athymic mice. These data link ROS production by Mox1 to growth control in non-phagocytic cells.

A human expressed sequence tag (EST) sequence which showed homology to human gp91phox was identified, and complete sequencing revealed a predicted amino-acid sequence homologous to the carboxy-terminal half of gp91phox. Sequencing was completed with 5'-rapid amplification of complementary DNA ends (5'-RACE) using human colon cDNA, and the predicted amino-acid sequence is shown in Fig. 1. The predicted protein is 564 amino acids long, compared with 569 residues for gp91phox, and the two show 56% identity. The gene is located at Xq22 (Genebank accession no. Z83819), but the locus is not informative with regard to known diseases. Rat *mox1*, cloned by using degenerate primers based on sequences that are highly conserved between gp91phox and human *mox1*, is 82% identical to human *mox1* (Fig. 1). Human and rat Mox1 lack asparagine-linked candidate glycosylation sites, which are seen in the highly glycosylated human and mouse

gp91phox/human	1	MGNWAVNEGLSIFAILVWGLNVFLVYVYVYDIPPKFFYTRKLGLSAL	50
p65mox/human		MGNWVNVHWFVFLVFLVWGLNVFLVDAFLKYEKADKYVYTRKILGSTL	
p65mox/rat		MGNWLNVHWSLVFLVFLVWGLNIFLVYVFLNYSKDYKYVYTRILGTAL	
gp91phox/human	51	ALARAPAAACLNFCMLILLPVCNRLLSFLRGSSACCSTRVRQLDRNLTF	100
p65mox/human		ACARASALCLNFNSTLILLPVCNRLLSFLRGTCSCFCSRTLRLKLDHNLTF	
p65mox/rat		ALARASALCLNFNSMVLIPVCNRLLSFLRGTCSCFNHTRLRLKLDHNLTF	
gp91phox/human	• • • • •	HKMVAWMIALHSAHTIAHLFNVEVCNARVNSDPYSVALSEL . . GDRQ	148
p65mox/human		HKLVAVMICLHTAHTIAHLFNFDYCSRSQATDGSALSSLSHDEKK	
p65mox/rat		HKLVAVMICIFTAHTIAHLFNFERYSRSQAMDGSALSVLSLHFHPEKE	
gp91phox/human	†	NESYLFNFARKRIKNEPGL . YLAVTLTLAGITGVVITLCLILITSSTKTI	197
p65mox/human		GGSWLN . . . PIQSRNTTVEYVFTSVAGLTGVIMTIALIMVTSATEFI	
p65mox/rat		D . SWLN . . . PIQSPNVTVMYAAFTSIAGLTGVVATVAVLVMVTSAMEFI	
gp91phox/human	198	RRSYFEVFWYTHHLFVIFFIGLAIHGAERIVRGQTAEASLAVHNITVCEQK	247
p65mox/human		RRSYFEVFWYTHHLFVIFFIGLAIHGAERIVRGQTAEASLAVHNITVCEQK	
p65mox/rat		RRNYFELFWYTHHLFVIFFIGLAIHGAERIVRGQTAEASLAVHNITVCEQK	
gp91phox/human	248	ISEWGGKIK . ECIPIQFAGNPMTWKVIGVPMFLYLCERLVFRWRSQQKVV	296
p65mox/human		FEMWDDRDSDHCRPKEFGHPPESSWKWILAPVILYICERILRFYRSQQKVV	
p65mox/rat		FHEWDKYSERCSRPFVGVQPPESWKWILAPVILYIFERILRFYRSQQKVV	
gp91phox/human	297	ITKVVTHPFKTIQLQMKKKGKMEVGYIFVCKPKVSKLEWHFPFTLSAP	346
p65mox/human		ITKVVTHPFKTIQLQMKKKGKMEVGYIFVCKPKVSKLEWHFPFTLSAP	
p65mox/rat		ITKVVTHPFKTIQLQMKKKGKMEVGYIFVCKPKVSKLEWHFPFTLSAP	
gp91phox/human	347	EEDFFSIHIRIVGDWTEGLFNACGCDKQEFQDAWK . LPKIADVDPFGTAS	395
p65mox/human		EEDFFSIHIRIVGDWTEGLFNACGCDKQEFQDAWK . LPKIADVDPFGTAS	
p65mox/rat		EEEEESIHIRIVGDWTEGLFNACGCDKQEFQDAWK . LPKIADVDPFGTAS	
gp91phox/human	396	EDVFSYEVVLMVAGIGVTPFASILKSWVYKYNATNLKLLKKIYFYWLC	445
p65mox/human		EDVFSYEVVLMVAGIGVTPFASILKSWVYKYNATNLKLLKKIYFYWLC	
p65mox/rat		EDVFSYEVVLMVAGIGVTPFASILKSWVYKYNATNLKLLKKIYFYWLC	
gp91phox/human	446	RDTHAFEFADLLQLLESQMQRNAGFLSYNIYLTGWDSQANHFVHH	495
p65mox/human		RETGAFAWNNLLNSLEQEMDELGKPDFLNRYLFLTGWDSNIVGHAALNF	
p65mox/rat		RETGAFAWNNLLNSLEQEMDELGKPDFLNRYLFLTGWDSNIVGHAALNF	
gp91phox/human	496	DEEKDVITGLKQKTLGYRPNWDNEFKTASQHPNTRIGVFLCGPEALAE	545
p65mox/human		DKATDITVGLKQKTLGYRPNWDNEFKTASQHPNTRIGVFLCGPEALAE	
p65mox/rat		DRATDITVGLKQKTLGYRPNWDNEFKTASQHPNTRIGVFLCGPEALAE	
gp91phox/human	546	LSKQSIINSSESGPRGVHIFNKENF	
p65mox/human		LRKCCRRYSSLDPRKVQFYFNKENF	
p65mox/rat		LRKCCRRYSSLDPRKVQFYFNKENF	

Figure 1 Predicted amino-acid sequence of human Mox1 (GenBank accession no. AF127763, patent pending) and rat Mox1 (GenBank accession no. AF152963), with human gp91phox. Abbreviations of amino-acid residues are according to standard IUPAC nomenclature. † and ‡ indicate asparagine-linked candidate glycosylation sites in mouse and human gp91phox, respectively. Filled circles indicate conserved histidines that are candidates for haem ligation.

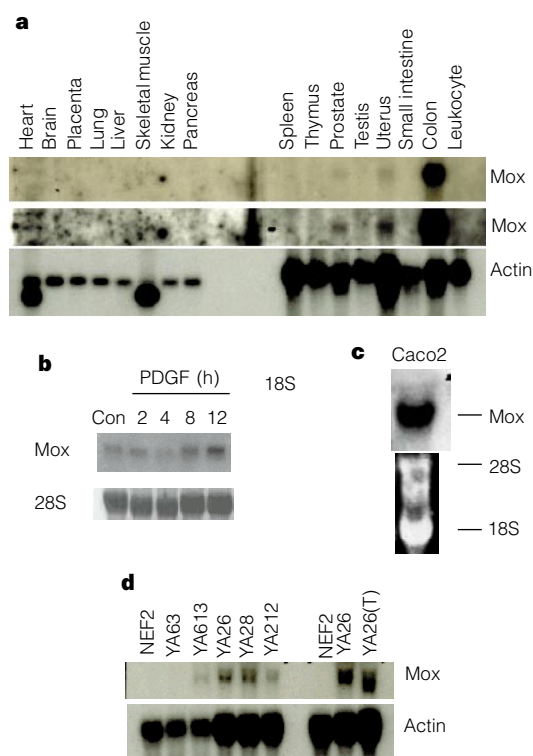


Figure 2 Expression of *mox1* mRNA. **a**, Tissue expression of human *mox1* mRNA. Top and middle panels were exposed for 24 h and 3 days, respectively. The bottom panel was probed for actin mRNA. **b**, Expression of rat *mox1* mRNA in rat aortic smooth muscle and induction by 20 ng ml⁻¹ PDGF. Bottom, 28S rRNA. **c**, Expression of human *mox1* in Caco-2 cells. **d**, Northern analysis of human *mox1* mRNA in NIH 3T3 cells transfected with vector alone (NEF2) or with human *mox1* (YA63, YA613, YA26, YA28 and YA212), and a tumour YA26(T) produced in athymic mice after injection of *mox1*-expressing cells.

gp91phox^{19,21} (the positions of which are indicated by ‡ and †, respectively, in Fig. 1). Regions in gp91phox (underlined), which were previously proposed^{21–22} to be binding sites for flavin (regions 1a and 1b) and pyridine nucleotide (regions 2a–2d), are identical or nearly identical between gp91phox and Mox1. Also shown (filled circles) are conserved histidines, which are candidates for haem ligation. The hydropathy profiles of human gp91phox and Mox1 are nearly identical (not shown), and include five very hydrophobic stretches in the amino-terminal half of the molecules which are predicted to be membrane-spanning regions.

The distribution in the tissues of *mox1* mRNA was very different from that of gp91phox. Among normal human tissues, *mox1* was expressed most in colon, and low expression was also detected in the uterus and prostate (Fig. 2a). The human colon-carcinoma cell line Caco-2 also expressed large quantities of *mox1* message (Fig. 2c). In cultured rat aortic vascular smooth-muscle cells (Fig. 2b), platelet-derived growth factor (PDGF), which causes cell proliferation^{13,24}, induced a roughly twofold increase in the expression of rat *mox1* mRNA within 12 hours of treatment, consistent with the idea that Mox1 contributes to the growth-stimulatory effects of PDGF.

To investigate whether Mox1 catalysed superoxide generation in cells, *mox1* DNA was stably transfected into NIH 3T3 cells. After selection, 30 surviving colonies were screened by the polymerase chain reaction with reverse transcription (RT-PCR) for expression of *mox1* mRNA (not shown), and 22 of these showed expression of *mox1* message. Five colonies were selected for further analysis, of which four (YA613, YA26, YA28 and YA212) showed significant expression of *mox1* message (Fig. 2d). Superoxide generation by intact cells was monitored by using superoxide dismutase-sensitive reduction of nitroblue tetrazolium (Fig. 3a). YA26 and YA28 cells

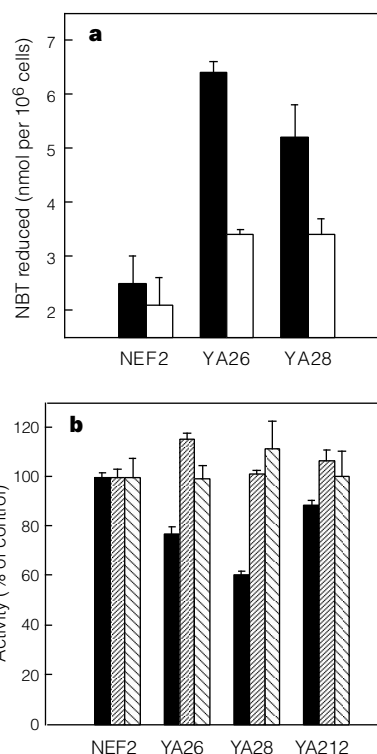


Figure 3 Superoxide generation by Mox1. **a**, Reduction of nitroblue tetrazolium (NBT) in *mox1*-transfected and control fibroblasts was measured in the absence (filled bars) or presence (open bars) of superoxide dismutase. **b**, Aconitase (filled bars), lactate dehydrogenase (narrow hatching) and fumarase (broad hatching) were determined in lysates of cells transfected with vector alone (NEF2) or with *mox1* (YA26, YA28 and YA212). Results are the average $\pm$ s.e.m. of 3 experiments. Activities are expressed as percentage of that seen in NEF2 cells.

showed increased reduction of nitroblue tetrazolium compared with vector (NEF2)-transfected cells, much of which was inhibited by superoxide dismutase. Because superoxide dismutase and nitroblue tetrazolium are not likely to penetrate cells, superoxide must be generated extracellularly. The amount of superoxide generated by these cells is on the order of 0.3–0.6 nmol min⁻¹ per 10⁶ cells, or about 5–10% of that generated by activated human neutrophils. NADPH-dependent superoxide generation was measured in broken cell preparations by using a lucigenin luminescence assay. Although substantial background signal was seen, luminescence was about twofold higher in *mox1*-transfected cells than in vector-transfected cells, and was inhibited by both superoxide dismutase and the general flavoprotein inhibitor diphenylene iodonium (not shown). This signal was unaffected by added recombinant human p47phox, p67phox and Rac1(GTP- γ S), which are essential cytosolic factors for the phagocyte respiratory-burst oxidase. Thus, *mox1*-transfected cells show increased superoxide generation compared with control cells.

To test whether superoxide generated by Mox1 can affect intracellular 'targets', we monitored aconitase activity in control and in three *mox1*-transfected cell lines. Aconitase contains a four-iron-four-sulphur cluster, which is highly susceptible to modification by superoxide, resulting in a loss in enzyme activity^{25,26}, and has been used as a reporter of intracellular superoxide generation. Aconitase activity was significantly diminished in YA26, YA28 and YA212 cells compared with the transfected control (Fig. 3b), and the decrease in activity paralleled the expression of *mox1* in these cells (compare with Fig. 2d). Approximately 50% of the aconitase in these cells is mitochondrial, based on differential centrifugation, and the cytosolic and mitochondrial forms were both affected (not shown).

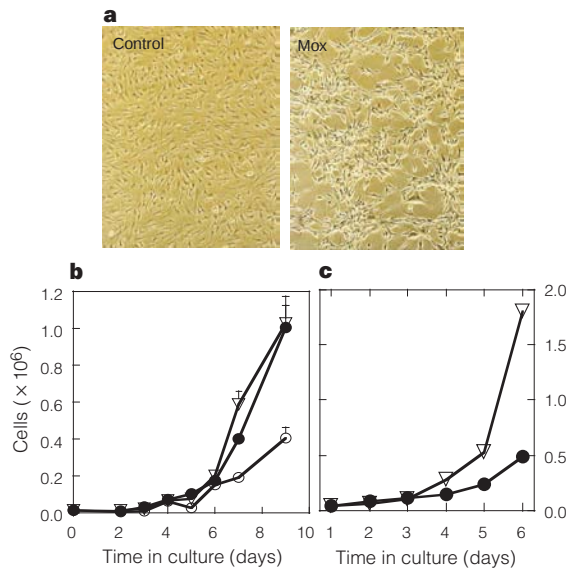


Figure 4 Properties of NIH 3T3 cells transfected with *mox1*. **a**, Control NIH 3T3 (left) and YA28 cells stably transfected with *mox1* (right) grown in culture for 3 days and observed by light microscopy (original magnification $\times 100$). **b**, Growth curves for vector control (NEF2) cell line (open circles) and YA26 (filled circles) and YA28 (open triangles) *mox1*-transfected cell lines. Cells were plated at 13×10^4 cells per well in 6-well plates. Triplicates $\pm$ s.e.m. are shown. **c**, Growth curves for YA28 cells in the absence (open triangles) or presence (filled circles) of 20 mM N-acetyl cysteine added daily to cultures.

Control cytosolic and mitochondrial enzymes that do not contain iron-sulphur centres were not affected (Fig. 3b). Superoxide generated in *mox1*-transfected cells is therefore capable of reacting with and modifying intracellular components.

Remarkably, *mox1*-transfected NIH 3T3 cells grown in culture take on a transformed appearance, becoming elongated and losing contact inhibition (Fig. 4a). Of 13 cell lines developed from *mox1*-transfected cells, 11 (85%) showed a transformed phenotype, whereas none of 22 lines (0%) isolated from cells transfected with vector alone showed this phenotype. The *mox1* cells retain a partial dependence on serum for growth (not shown), and growth rates in the presence of serum are increased (Fig. 4b). Growth of the *mox1*-transfected cells was inhibited by the antioxidant N-acetyl cysteine (Fig. 4c), consistent with reactive oxygen acting as mediator in the regulation of cell division. The *mox1*-transfected cells exhibited anchorage-independent growth in soft agar (not shown), although colonies were smaller than those seen for *Ras*-transformed fibroblasts.

The *mox1*-transfected cells also produced aggressive tumours in athymic mice (not shown), which at about 3–4 weeks were similar in size to those produced by *ras*-transformed NIH 3T3 cells. Of 15 mice injected with *mox1*-transfected NIH 3T3 cells, 14 showed large tumours within 17 days of injection, and tumours showed expression of *mox1* mRNA (Fig. 2d). Histologically, the tumours resembled fibrosarcomas and were similar to *ras*-induced tumours (not shown). Thus, *ras* and *mox1* were similarly potent in their ability to induce tumorigenicity of NIH3T3 cells in athymic mice.

A role in normal growth was demonstrated in rat aortic vascular smooth-muscle cells by using antisense to rat *mox1*. Transfection with the antisense DNA resulted in a decrease in both superoxide generation (Fig. 5b) and serum-dependent growth (Fig. 5a). *Mox1* is therefore implicated in normal growth in this cell type.

Our results show that *Mox1*, a homologue of gp91phox, generates ROS in some non-phagocytic cells and may participate in normal mitogenic regulation in response to growth factors, such as PDGF, and in dysregulated growth in hyperproliferative disorders such as cancer and atherosclerosis. □

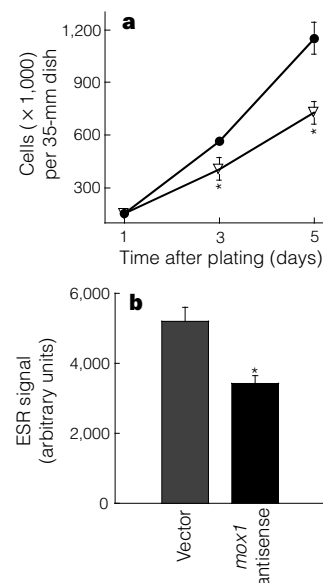


Figure 5 Role of *Mox1* in growth of vascular smooth-muscle cells. **a**, Cells were transfected with antisense rat *mox1* (open triangles) or empty vector (filled circles), and cell number was quantified using a Coulter counter. Values are the mean $\pm$ s.e.m. of 3 experiments; $P < 0.001$. **b**, Cells were transfected with antisense rat *mox1* or empty vector and were analysed for superoxide by electron spin resonance (ESR). Bars show the mean $\pm$ s.e.m. of 3 independent experiments. $P = 0.03$.

Methods

Cloning.

A homologous 334-base portion of an expressed sequence tag (EST176696; GenBank accession no. AA305700) from a Caco-2 cell line was identified by using the human gp91phox amino-acid sequence as a query in a Blast search. The bacterial strain #129134 containing the EST sequence in the pBluescript SK-vector was purchased from ATCC (Rockville). EST176696 encoded 337 amino acids showing homology to the C-terminal 60% of human gp91phox, and contained a stop codon corresponding to the C terminus of gp91phox. 5'-RACE, carried out using a human colon cDNA library and Marathon cDNA Amplification Kit (Clontech, Palo Alto) using a gene-specific primer, yielded a 1.1-kilobase (kb) fragment containing an ATG codon corresponding in position to the start site in gp91phox. Reamplification with DNA primers spanning the putative start and stop codons yielded a 1.7-kb fragment which was TA-cloned into the PCR2.1 vector (Invitrogen TA Cloning Kit, San Diego).

A 1.1-kb central fragment of rat *mox1* cDNA was obtained by RT-PCR of vascular smooth-muscle-cell RNA using primers designed to anneal to human *mox1* and human, mouse and pig gp91phox. The sequence was extended by 5'- and 3'-RACE, which yielded 850-base pair and 1.5-kb products, respectively. Full-length rat *mox1* cDNA (2.6 kb) was obtained by RT-PCR of the smooth-muscle-cell RNA by using primers designed to anneal to the terminal segments of the 5'- and 3'-RACE products.

Transfections.

PEF-PAC vector alone and vector containing the subcloned coding region for *mox1* were stably transfected ($10 \mu\text{g}$ DNA) into 2×10^6 NIH 3T3 cells by using Fugene 6 (Boehringer Mannheim). Cells maintained in DMEM containing 10% calf serum were split after 2 days and selected with $1 \mu\text{g ml}^{-1}$ puromycin. Colonies surviving in selection media for 10–14 days were subcultured and characterized for expression of *mox1* mRNA. For antisense experiments, cells plated at low density were transfected using Superfect reagent (Qiagen) with either full-length antisense rat-*mox1* in the pcDNA3 vector (Invitrogen) or empty vector. Cells were replated at higher density, allowed to grow in medium with 10% calf serum, $400 \mu\text{g ml}^{-1}$ geneticin and 10 mM glutamine and passaged 3–4 times in selection medium. Transcription from the empty vector and the antisense construct were verified by RT-PCR by using vector-specific primers.

Northern blots.

Human multiple tissue northern blot I and blot IV membranes (Clontech) were incubated at 68°C in hybridization solution for 1 h and hybridized (68°C , 3–4 h) with radiolabelled *mox1* coding region. Human or rat *mox1* coding region was labelled by random priming with [^{32}P]dCTP ($10 \mu\text{Ci}$) using the Prime-It II kit (Stratagene, La Jolla). For analysis of *mox1* mRNA expression in cell lines, total RNA was prepared from 10^6 cells using the High Pure RNA Isolation Kit (Boehringer Mannheim) or RNeasy kit (Qiagen, Valencia, CA). Total RNA ($10\text{--}20 \mu\text{g}$) was separated on a 1% agarose formaldehyde minigel and transferred to a Nytran filter (Biorad, Hercules, CA) and immobilized by ultraviolet crosslinking.

Nitroblue tetrazolium (NBT) reduction assay.

Cells (500,000 per well in six-well plates) were plated. After 24 h, medium was removed and cells were washed with 1 ml Hanks solution (Sigma). Filtered 0.25% NBT (1 ml, Sigma) was added with or without 600 units of superoxide dismutase (Sigma) and cells were incubated at 37 °C for 8 min. Cells were scraped and NBT was pelleted by low-speed centrifugation. The NBT pellet was resuspended in 1 ml of pyridine (Sigma) and solubilized by heating for 10 min at 100 °C. Reduced NBT was quantified by measuring the absorbance at 510 nm (extinction coefficient of 11,000 M⁻¹ cm⁻¹).

Aconitase, lactate dehydrogenase and fumarase assays.

Aconitase, lactate dehydrogenase and fumarase activities were determined as described^{25,27,28}.

Superoxide measurement by spin trapping.

Transfected cells were washed twice with phosphate-buffered saline (PBS), sedimented at 400g for 10 min, resuspended and homogenized by sonication in 1 ml lysis buffer (50 mM Tris, 0.1 mM EDTA, 0.1 mM EGTA, aprotinin, leupeptin, pepstatin and PMSF). Membrane pellet (10 µg 28,000g) (resuspended in lysis buffer) was incubated with 10 mM DEPMPO (5-diethoxyphosphoryl-5-1-pyrroline-N-oxide, Oxis), 100 µM diethyldithiocarbamate, 100 µM NADPH, 0.1 mM diethylenetriaminepentaacetic acid in PBS at 37 °C for 30 min. Electron spin resonance spectra were recorded using a Bruker EMX spectrometer at 9.78 GHz, 20 mW with a modulation amplitude of 1.0 G, and quantified by adding the total height of eight peaks relative to baseline.

Superoxide measurement by lucigenin.

Lucigenin luminescence measurements were carried out using 30 µM lucigenin²⁹. When added, p47phox, p67phox and Rac1(GTP-γS) were at 1 µM.

Soft agar assay.

Trypsinized cells were suspended in medium containing DMEM, 10% fetal bovine serum, 100 units ml⁻¹ penicillin, 100 µg ml⁻¹ streptomycin and 0.3% noble agar (Difco, Detroit) with or without puromycin; 1.5 × 10⁵ cells per well were plated in six-well plates onto solidified medium containing 1% noble agar. Plates were incubated at 37 °C in 5% CO₂, and medium was added every 5–7 days. After 19 days, cells were incubated overnight with p-iodonitrotetrazolium violet (Sigma) and photographed.

Received 21 May; accepted 16 July 1999.

- Jones, S. A., Hancock, J., Jones, O. T. G., Neubauer, A. & Topley, N. The expression of NADPH oxidase components in human glomerular mesangial cells: Detection of protein and mRNA for p47phox, p67phox, and p22phox. *J. Am. Soc. Nephrol.* **5**, 1483–1491 (1995).
- Matsubara, T. & Ziff, M. Increased superoxide anion release from human endothelial cells in response to cytokines. *J. Immunol.* **137**, 3295–3298 (1986).
- Meier, B. *et al.* Human fibroblasts release reactive oxygen species in response to interleukin-1 or tumor necrosis factor-α. *Biochem. J.* **263**, 539–545 (1989).
- Pagano, P. *et al.* Localization of a constitutively active, phagocyte-like NADPH oxidase in rabbit aortic adventitia: Enhancement by angiotensin II. *Proc. Natl Acad. Sci. USA* **94**, 14483–14488 (1997).
- Jones, S. *et al.* Expression of phagocyte NADPH oxidase components in human endothelial cells. *Am. J. Physiol.* **271**, H1626 (1996).
- Griendling, K. K., Minieri, C. A., Ollerenshaw, J. D. & Alexander, R. W. Angiotensin II stimulates NADH and NADPH oxidase activity in cultured vascular smooth muscle cells. *Circ. Res.* **74**, 1141–1148 (1994).
- Szattowski, T. P. & Nathan, C. F. Production of large amounts of hydrogen peroxide by human tumor cells. *Cancer Res.* **51**, 794–798 (1991).
- Burdon, R. Superoxide and hydrogen peroxide in relation to mammalian cell proliferation. *Free Rad. Biol. Med.* **18**, 775–794 (1995).
- Crawford, D., Zbinden, L., Amstad, P. & Cerutti, P. Oxidant stress induces the proto-oncogenes *c-fos* and *c-myc* in mouse epidermal cells. *Oncogene* **3**, 27–32 (1988).
- Nose, K. *et al.* Transcriptional activation of early-response genes by hydrogen peroxide in a mouse osteoblastic cell line. *Eur. J. Biochem.* **201**, 99–106 (1991).
- Datta, R. *et al.* Involvement of reactive oxygen intermediates in the induction of c-jun gene transcription by ionizing radiation. *Biochemistry* **31**, 8300–8306 (1992).
- Bae, Y. S. *et al.* Epidermal growth factor (EGF)-induced generation of hydrogen peroxide. *J. Biol. Chem.* **272**, 217–221 (1997).
- Zafari, A. M. *et al.* Role of NADH/NADPH oxidase-derived H₂O₂ in angiotensin II-induced vascular hypertrophy. *Hypertension* **32**, 488–495 (1998).
- Wolin, M. S. Activated oxygen metabolites as regulators of vascular tone. *Klin. Wochenschr.* **69**, 1046–1049 (1991).
- Sundaresan, M., Zu-Xi, Y., Ferrans, V. J., Irani, K. & Finkel, T. Requirement for generation of H₂O₂ for platelet-derived growth factor signal transduction. *Science* **270**, 296–299 (1995).
- Irani, K. *et al.* Mitogenic signaling mediated by oxidants in Ras-transformed fibroblasts. *Science* **275**, 1649–1652 (1997).
- Segal, A. W. & Shatwell, K. P. The NADPH oxidase of phagocytic leukocytes. *Ann. NY Acad. Sci.* **832**, 215–222 (1997).
- Royer-Pokora, B. *et al.* Cloning the gene for an inherited human disorder—chronic granulomatous disease—on the basis of its chromosomal location. *Nature* **322**, 32–38 (1986).
- Wallach, T. M. & Segal, A. W. Analysis of glycosylation sites on gp91phox, the flavocytochrome of the NADPH oxidase, by site-directed mutagenesis and translation in vitro. *Biochem. J.* **321**, 583–5746 (1998).
- Bjorgvinsdottir, H., Zhen, L. & Dinanier, M. Cloning of murine gp91phox cDNA and functional expression in a human X-linked chronic granulomatous disease cell line. *Blood* **87**, 2005–2010 (1996).
- Rotrosen, D., Yeung, C. L., Leto, T. L. & Malech, K. C. H. Cytochrome b₅₅₈: The flavin-binding

component of the phagocyte NADPH oxidase. *Science* **256**, 1459–1462 (1992).

- Segal, A. W. *et al.* Cytochrome b₂₄₅ is a flavocytochrome containing FAD and the NADPH-binding site of the microbicidal oxidase of phagocytes. *Biochem. J.* **284**, 781–788 (1992).
- Sumimoto, H. *et al.* Cytochrome b₅₅₈, a component of the phagocyte NADPH oxidase, is a flavoprotein. *Biochem. Biophys. Res. Commun.* **186**, 1368–1363 (1992).
- Berk, B. C., Elder, E. & Mitsuka, M. Hypertrophy and hyperplasia cause differing effects on vascular smooth muscle cell Na⁺/H⁺ exchange and intracellular pH. *J. Biol. Chem.* **265**, 19632–19637 (1990).
- Gardner, P., Rainieri, L., Epstein, L. & White, C. Superoxide radical and iron modulate aconitase activity in mammalian cells. *J. Biol. Chem.* **270**, 13399–13405 (1995).
- Fridovich, I. Superoxide anion radical (O₂⁻), superoxide dismutases, and related matters. *J. Biol. Chem.* **272**, 18515–18517 (1997).
- Gardner, P. R., Znguyen, D. H. & White, C. W. Aconitase is a sensitive and critical target of oxygen poisoning in cultured mammalian cells and in rat lungs. *Proc. Natl Acad. Sci. USA* **91**, 12248–12252 (1994).
- Flint, D. H. Initial kinetic and mechanistic characterization of *Escherichia coli* fumarase A. *Arch. Biochem. Biophys.* **311**, 509–513 (1994).
- Li, Y. *et al.* Validation of lucigenin (bis-N-methylacridinium) as a chemiluminescent probe for detecting superoxide anion radical production by enzymatic and cellular systems. *J. Biol. Chem.* **273**, 2015–2023 (1998).

Acknowledgements

We thank D. Jones and D. Edmondson for discussions and technical advice; W. Sun for technical assistance; and D. Dillehay for the histological analysis. This work was supported by the NIH.

Correspondence and requests for materials should be addressed to J.D.L. (e-mail: dlambe@bimcore.emory.edu).

NF-κB activation by tumour necrosis factor requires the Akt serine–threonine kinase

Osman Nidai Ozes*, Lindsey D. Mayo*, Jason A. Gustin*, Susan R. Pfeffer†, Lawrence M. Pfeffer† & David B. Donner*

* Department of Microbiology and Immunology, Indiana University School of Medicine and the Walther Oncology Center, Indianapolis, Indiana 46202, USA

† Department of Pathology, University of Tennessee Health Science Center, Memphis, Tennessee 38163, USA

Activation of the nuclear transcription factor NF-κB by inflammatory cytokines requires the successive action of NF-κB-inducing kinase (NIK) and an IκB-kinase (IKK) complex composed of IKKα and IKKβ^{1–5}. Here we show that the Akt serine–threonine kinase⁶ is involved in the activation of NF-κB by tumour necrosis factor (TNF). TNF activates phosphatidylinositol-3-OH kinase (PI(3)K) and its downstream target Akt (protein kinase B). Wortmannin (a PI(3)K inhibitor), dominant-negative PI(3)K or kinase-dead Akt inhibits TNF-mediated NF-κB activation. Constitutively active Akt induces NF-κB activity and this effect is blocked by dominant-negative NIK. Conversely, NIK activates NF-κB and this is blocked by kinase-dead Akt. Thus, both Akt and NIK are necessary for TNF activation of NF-κB. Akt mediates IKKα phosphorylation at threonine 23. Mutation of this amino acid blocks phosphorylation by Akt or TNF and activation of NF-κB. These findings indicate that Akt is part of a signalling pathway that is necessary for inducing key immune and inflammatory responses.

NF-κB regulates gene expression in immunity, stress responses, inflammation⁷ and the inhibition of apoptosis^{8,9}. Ordinarily, NF-κB is sequestered in the cytoplasm by the inhibitory protein IκB. Ultraviolet radiation, T-cell activation, bacterial lipopolysaccharide, viral gene products and inflammatory cytokines promote IκB degradation, thereby allowing NF-κB to enter the nucleus and induce gene transcription⁷. Activation of the type-1 TNF receptor

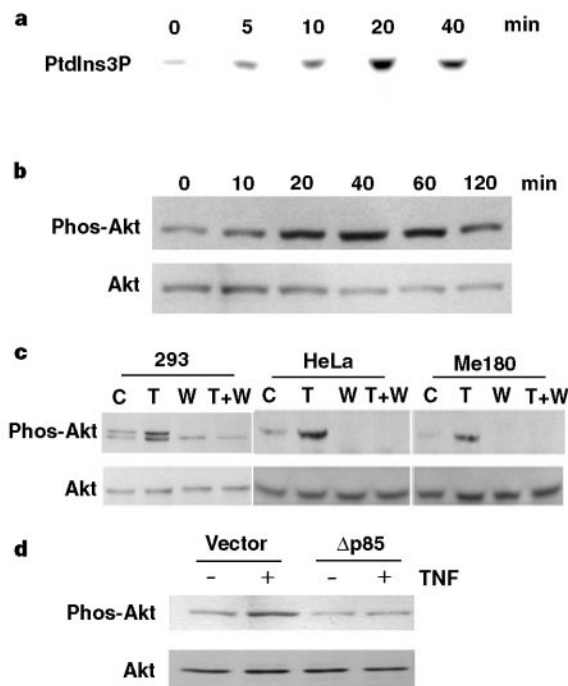


Figure 1 Activation of PI(3)K and Akt by TNF. Serum-starved (16 h) 293 cells were incubated with 1 nM TNF at 37 °C. **a**, Endogenous PI(3)K activity in p85 immunoprecipitates measured by lipid-kinase assays²⁰. After 5, 10, 20 and 40 min of treatment, activity increased 3.8 ± 0.9 ($P < 0.013$), 5.0 ± 0.8 ($P < 0.035$), 7.2 ± 1.0 ($P < 0.006$) and 7.02 ± 2.2 ($P < 0.027$) fold, respectively. **b**, Endogenous Akt activity assayed by western blotting. Relative Akt phosphorylation was increased by TNF 1.7 ± 0.3 ($P < 0.01$), 2.2 ± 0.2 ($P < 0.005$) and 1.5 ± 0.3 ($P < 0.04$) fold, at 20, 40 and 60 min, respectively. **c**, 293, HeLa or ME-180 cells were incubated with or without 100 nM wortmannin for 30 min and then with TNF (30 min). A western blot of endogenous proteins was probed with antibodies to phospho-Akt and then Akt. **d**, TNF (1 nM, 30 min) activation of Akt in cells transfected with empty vector or Δ p85 assayed by western blotting.

(TNFR1) induces the formation of a partially characterized signalling complex¹⁰ that contains TNF-receptor-associated-factor 2 (TRAF2)¹¹, which binds NIK, a MAP kinase kinase kinase¹². Phosphorylation of the IKK α component of the I κ B kinase complex by NIK targets I κ B for degradation and induces NF- κ B activation¹³.

TNF activates TNFR1 and thereby induces the tyrosine phosphorylation of the p85 subunit of PI(3)K¹⁴, leading us to directly test the effect of TNF on PI(3)K activity. In embryonic kidney 293 cells (Fig. 1a) and ME-180 human cervical carcinoma cells (data not shown), the lipid-kinase activity of PI(3)K was increased within 5 min by TNF, and maximal induction occurred after 20 min. As Akt is a downstream substrate for PI(3)-kinase⁶, we tested whether Akt is involved in cellular responses to TNF by probing western blots of proteins from 293 cells with antibodies to phosphorylated (activated) Akt (Fig. 1b, top) and Akt (Fig. 1b, bottom). We found that TNF induced a time-dependent increase in Akt phosphorylation, which was temporally correlated with PI(3)K activation (Fig. 1a). TNF also activated Akt in HeLa and ME-180 cells, and this response

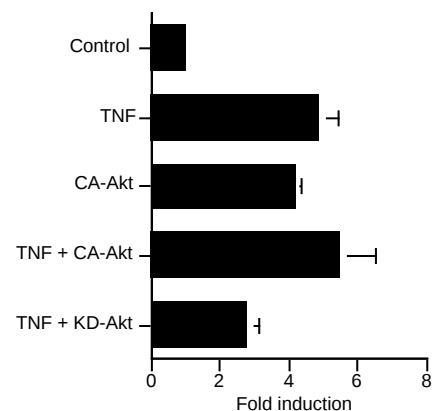


Figure 3 NF- κ B reporter activity in cells transfected with empty vector, CA-AKT or KD-Akt and then treated with vehicle or TNF. Results are the average of three experiments conducted in triplicate $\pm$ s.d.

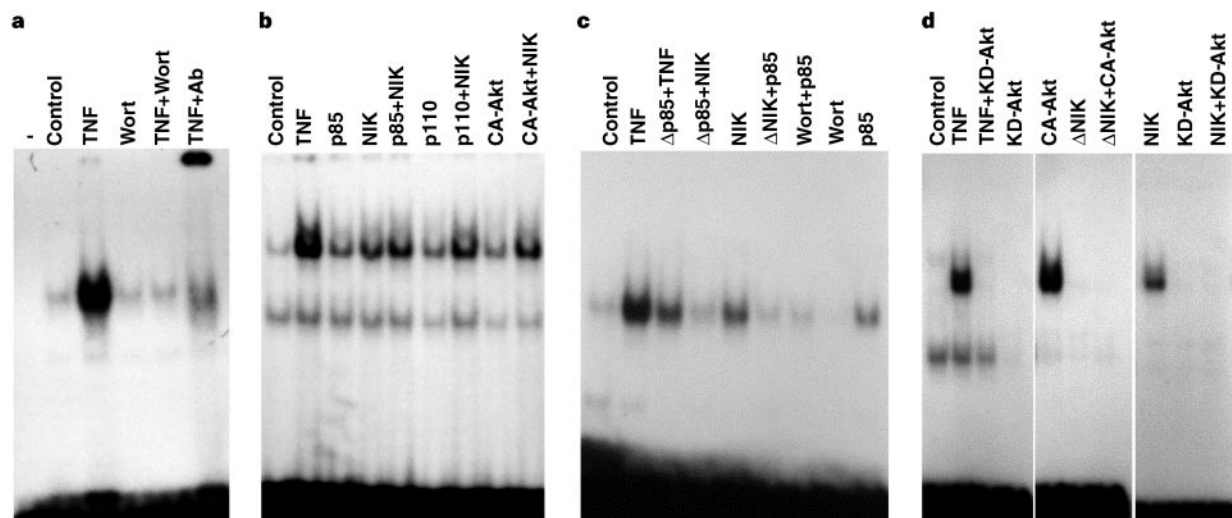


Figure 2 NF- κ B activation by a TNF-stimulated PI(3)K–Akt pathway. For EMSA (**a–d**), control lanes contain extracts from unstimulated cells (control) or free probe (–). Supershifts by anti-Rel A showed that the retarded band was an NF- κ B–DNA complex. **a**, Cells were pretreated with vehicle or 100 nM wortmannin for 30 min before incubation with 1 nM TNF (30 min). Cells were transfected with: **b**, p85, NIK, p110, CA-AKT or control

DNA. A sample transfected with control DNA was treated with TNF for reference; **c**, p85, NIK, Δ NIK, Δ p85 or control DNA. Transfected cells were treated with TNF or wortmannin as indicated in **a**; **d**, left, control DNA or KD-Akt before treatment with vehicle or TNF; middle, CA-Akt, Δ NIK or CA-Akt and Δ NIK; right, NIK, KD-Akt or NIK and KD-Akt. Wort, wortmannin; Ab, antibody.

was blocked by wortmannin (Fig. 1c). To determine whether Akt is downstream of PI(3)K in TNF signalling, 293 cells were transiently transfected with dominant-negative PI(3)K ($\Delta p85$), which abrogated the 2.3-fold increase in Akt activity promoted by TNF (Fig. 1d).

We next tested whether the PI(3)K–Akt pathway would influence NF- κ B DNA-binding activity, using electrophoretic mobility shift assays (EMSA). TNF induced NF- κ B activity in 293 (Fig. 2a), HeLa and ME-180 cells (data not shown), and wortmannin inhibited this effect. NF- κ B DNA-binding activity was enhanced by transiently transfected, constitutively active PI(3)K (the p85 regulatory subunit or the p110 catalytic subunit) or Akt (CA-Akt), or NIK (Fig. 2b). Furthermore, co-transfection of NIK with p85, p110 or CA-Akt produced an additive effect on NF- κ B DNA-binding activity. In contrast, $\Delta p85$ inhibited TNF or NIK activation of NF- κ B, whereas dominant-negative NIK (Δ NIK) or wortmannin abrogated activation of NF- κ B by PI(3)K (Fig. 2c). These observations show that PI(3)K and Akt are components of a previously unrecognized signalling pathway leading to NF- κ B activation.

Kinase-dead Akt (KD-Akt) inhibited TNF-promoted DNA binding by NF- κ B (Fig. 2d, left) and CA-Akt by itself induced such binding (Fig. 2d, middle), establishing that Akt is essential for NF- κ B activation. The ability of CA-Akt to induce NF- κ B binding was inhibited by Δ NIK (Fig. 2d, middle), and NF- κ B activation by NIK was inhibited by KD-Akt (Fig. 2d, right). NF- κ B reporter assays confirmed the results from EMSA, showing that CA-Akt or TNF

increased NF- κ B-dependent gene activity (>4 -fold), whereas the effect of TNF was blocked by KD-Akt (Fig. 3). Thus, Akt and NIK are both necessary for NF- κ B activation.

We identified a putative Akt phosphorylation site¹⁵ at amino-acids 18–23 in IKK α , which led us to test whether IKK α is a substrate for Akt. Akt and IKK α constitutively associate *in vivo*, and this interaction is unaffected by TNF or wortmannin (Fig. 4a). Immunodepletion showed that 12% of the cellular IKK α was associated with Akt (data not shown). TNF increased Akt phosphorylation in IKK α immunoprecipitates (3.3 ± 0.1 fold, $P < 0.023$, $n = 3$) (Fig. 4b) and Akt kinase activity (3.2 ± 0.4 fold, $P < 0.001$), effects that are inhibited by wortmannin (Fig. 4c).

Akt increased IKK α phosphorylation significantly (Fig. 5a, b), whereas only minor phosphorylation was detected on an IKK α mutant (T23A) in which threonine 23 in the putative Akt phosphorylation site of IKK α was mutated to alanine (T23A). This may be attributed to other kinases in Akt immunoprecipitates. Addi-

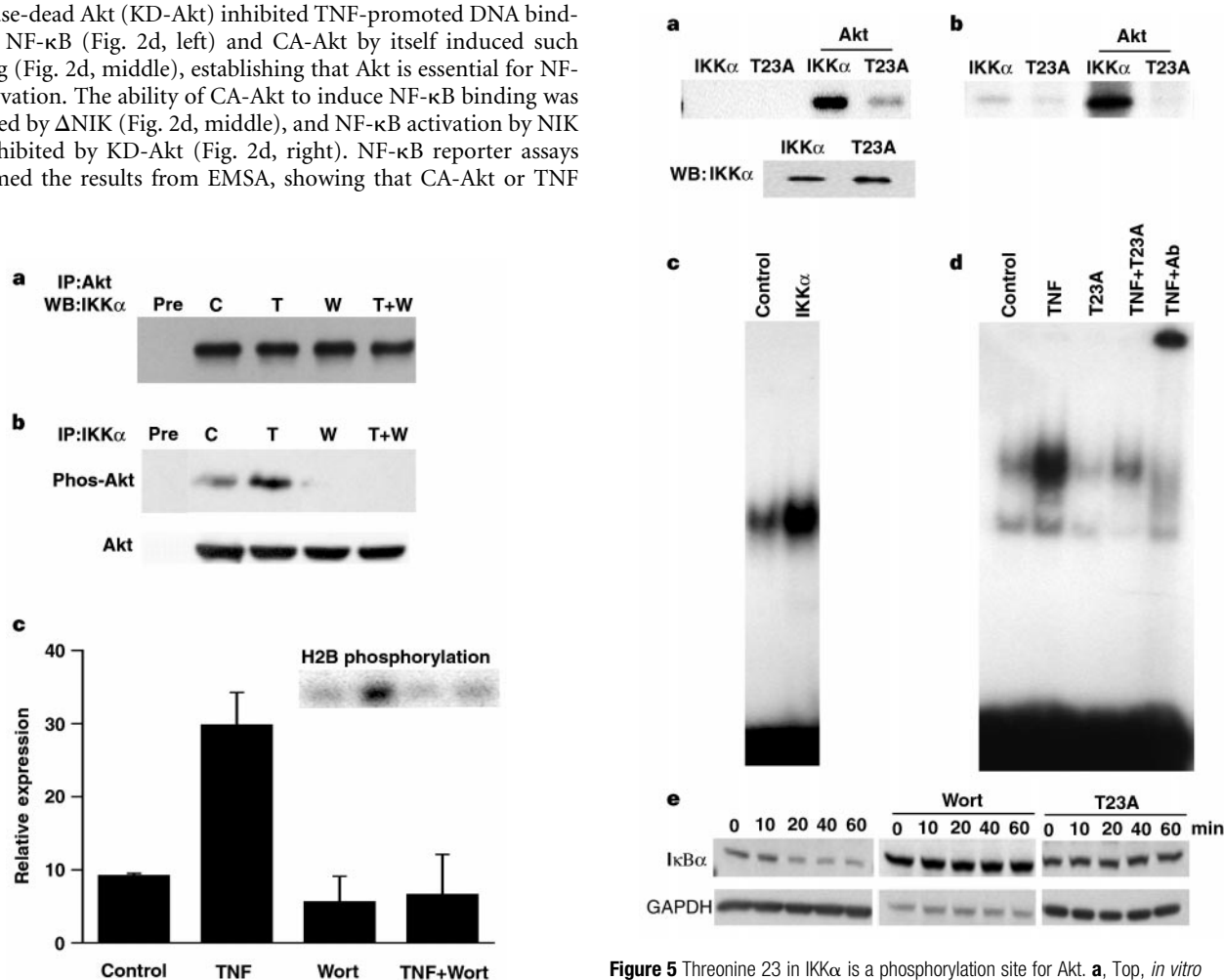


Figure 5 Threonine 23 in IKK α is a phosphorylation site for Akt. **a**, Top, *in vitro* phosphorylation of equal amounts of GST-IKK α or GST-T23A incubated in the absence or presence of activated Akt. Bottom, a western blot of each GST fusion protein used in the phosphorylation reaction was probed with an antibody to IKK α . **b**, *In vitro* phosphorylation reactions in the absence or presence of activated Akt performed on transiently expressed IKK α and T23A immunoprecipitated from 293 cells. **c**, EMSA of 293 cells transiently expressing IKK α or **d**) T23A incubated with vehicle or TNF. **e**, Top, cells were transfected with control DNA (left) or T23A (right) and after 24 h were incubated in serum-free medium (24 h). Cycloheximide ($50 \mu\text{g ml}^{-1}$) was added; 1 h thereafter the cells were treated with TNF before assay of I κ B expression by western blot analysis. Middle, serum-starved cells incubated with 100 nM wortmannin and $50 \mu\text{g ml}^{-1}$ cycloheximide for 1 h and then with TNF were assayed for I κ B. Bottom, blots were stripped and probed with an antibody to GAPDH.

Figure 4 Association of Akt with IKK α . Serum-starved 293 cells were pretreated with vehicle or wortmannin (100 nM, 30 min) and then with TNF (1 nM, 30 min) before immunoprecipitation of endogenous proteins or *in vitro* kinase assays. **a**, Western blot of proteins immunoprecipitated by pre-immune serum or anti-Akt probed with anti-IKK α . **b**, Western blot of an anti-IKK α immunoprecipitate probed with antibodies to active (top) or total (bottom) Akt. **c**, Akt kinase activity in IKK α immunoprecipitates. Immunoprecipitated IKK α /Akt was resuspended in 50 mM Tris, pH 7.5, 10 mM MgCl₂ and 1 mM dithiothreitol, 2.5 μg histone H2B, 50 μM ATP, and 10 μCi [γ -³²P]ATP and reaction proceeded for 30 min at 30 °C. Data from four independent experiments was averaged and are presented $\pm$ s.e.m. (bar graph). Insert, a representative assay. C, control; T, TNF; W, wortmannin; T + W, TNF and wortmannin.

tionally, recombinant IKK α phosphorylated by active Akt *in vitro* or immunopurified from TNF-treated (30 min, 1 nM), serum-starved IKK α -transfected 293 cells was digested with trypsin and subjected to liquid chromatography–electron spray ionization–mass spectrometry. In both experiments, a chromatographic peak of a mass of 1,702 was observed with multiply charged ions (data not shown), corresponding to the calculated mass (1,701.8) of the phosphorylated peptide containing threonine 23 of the Akt phosphorylation site of IKK α (LGTGGFGNVCLYQHR). Thus, TNF mediates phosphorylation of IKK α at threonine 23 *in vivo*.

The functional significance of threonine-23 phosphorylation is illustrated by the observation that TNF activates NF- κ B in 293 cells expressing IKK α (Fig. 5c), but not in cells expressing T23A (Fig. 5d). Furthermore, TNF induced I κ B degradation within 20 min (Fig. 5e, left), wortmannin (Fig. 5e, middle) or T23A expression (Fig. 5e, right) inhibited such TNF-induced degradation of I κ B. These observations show that the PI(3)K–Akt pathway mediates TNF-promoted NF- κ B activation by phosphorylation of threonine 23 in IKK α .

In cells IKK α is complexed with IKK β , and each kinase contributes to NF- κ B activation^{1–4}. NF- κ B activation by TNF is diminished in cells derived from IKK β or IKK α knockout mice^{16–18}, indicating that each kinase can partially compensate for the loss of the other^{16–18}. Mutation of IKK α at serine 176 (the phosphorylation site for NIK¹³) or threonine 23 (the phosphorylation site for Akt) abolishes kinase activity and NF- κ B activation. As IKK β does not appear to be phosphorylated by NIK¹³ or contain a putative Akt phosphorylation site, the activities of IKK α and IKK β are apparently differentially regulated. The different developmental defects in IKK-knockout mice^{16–18}, the indications that IKK α and IKK β may reciprocally regulate one another^{16,19} and the distinct mechanisms through which the kinases are regulated indicate that each kinase has a different role in IKK function¹⁶. Thus, IKK α and IKK β appear to play substantive but partially redundant roles in NF- κ B activation by TNF.

Akt is a mitogen-activated survival factor⁷. Our study shows that a signalling pathway culminating in phosphorylation of IKK α by Akt is necessary for activation of NF- κ B. The model in Fig. 6 illustrates the relationship of the PI(3)K–Akt and TRAF2–NIK pathways that converge on IKK α to promote NF- κ B activation at two sites: threonine 23 (Akt-mediated) and serine 176 (NIK-mediated)¹³. Our results define a mechanism through which Akt

may suppress apoptosis, and implicate Akt in cytokine-mediated immunity. □

Methods

Constructs.

CA-Akt and KD-Akt were gifts from R. Roth; p85 and Δ p85 were gifts from M. Kasuga; and p110 was from L. T. Williams. The constitutively active PI(3)K and Akt constructs promoted cell proliferation, whereas the dominant-negative constructs induced cell death. NIK, Δ NIK (NIK K429A, NIK K430A) and IKK α were gifts from H. Y. Song. RSV β Gal and the κ B probe were gifts from H. Nakshatri.

NF- κ B assays.

EMSA were conducted on 6 μ g of protein from 293 cells transiently transfected with pCMV expression constructs. NF- κ B DNA binding was assayed using a double-stranded ³²P-labelled κ B probe (5'-AGTTGAGGGGACTTCCAGG-3'). Transfection efficiencies of >90% were achieved using calcium phosphate precipitation. Twenty-four hours after transfection, cells were serum-starved for 24 h and then treated with TNF. For reporter gene assays, cells were transiently cotransfected with RSV β , β -galactosidase and pE-luc, a reporter plasmid containing 861 base pairs of the κ B-containing E-selectin promoter inserted upstream of the luciferase reporter plasmid pGL2. Forty-eight hours after transfection cells were incubated with vehicle or TNF (1 nM) for 6 h before assaying luciferase activity, which was normalized on the basis of β -gal expression.

Isolation of activated Akt and preparation of IKK α –GST fusion proteins.

3T3 cells overexpressing Akt (from M. E. Greenberg) were stimulated with 20 ng ml^{–1} PDGF for 10 min and activated Akt was purified by immunoprecipitation. GST fusion proteins were created by ligating *Sal*I- and *Nor*I-digested IKK α or T23A cDNA into the pGEX-4T3 plasmid.

Reproducibility of data.

Experiments were repeated at least three times with consistent results.

Received 6 May; accepted 14 June 1999.

- DiDonato, J. A., Hayakawa, M., Rothwarf, D. M., Zandi, E. & Karin, M. A cytokine-responsive I κ B kinase that activates the transcription factor NF- κ B. *Nature* **388**, 548–554 (1997).
- Regnier, C. H. *et al.* Identification and characterization of an I κ B kinase. *Cell* **90**, 373–383 (1997).
- Mercurio, F. *et al.* IKK-1 and IKK-2: cytokine-activated I κ B kinases essential for NF- κ B activation. *Science* **278**, 860–866 (1997).
- Woronicz, J. D., Gao, X., Rothe, M. & Goeddel, D. V. I κ B kinase- β : NF- κ B activation and complex formation with I κ B kinase- α and NIK. *Science* **278**, 866–869 (1997).
- Zandi, E., Rothwarf, D. M., Delhase, M., Hayakawa, M. & Karin, M. The I κ B kinase complex (IKK) contains two kinase subunits, IKK α and IKK β , necessary for I κ B phosphorylation and NF- κ B activation. *Cell* **91**, 243–252 (1997).
- Marte, B. M. & Downward, J. PKB/Akt: connecting phosphoinositide 3-kinase to cell survival and beyond. *Trends Biochem. Sci.* **22**, 355–358 (1997).
- Bauerle, P. A. & Baltimore, D. NF- κ B: Ten years after. *Cell* **87**, 13–20 (1996).
- Wang, C.-Y., Mayo, M. W. & Baldwin, A. S. Jr TNF- and cancer therapy-induced apoptosis: potentiation by inhibition of NF- κ B. *Science* **274**, 784–787 (1996).
- Van Antwerp, D. J., Martin, S. J., Kafri, T., Green, D. R. & Verma, I. M. Suppression of TNF- α -induced apoptosis by NF- κ B. *Science* **274**, 787–789 (1996).
- Maniatis, T. Catalysis by a multiprotein I κ B kinase complex. *Science* **278**, 818–819 (1997).
- Hsu, H., Shu, H.-B., Pan, M.-G. & Goeddel, D. V. TRADD–TRAF2 and TRADD–FADD interactions define two distinct TNF receptor 1 signal transduction pathways. *Cell* **84**, 299–308 (1996).
- Malinin, N. L., Boldin, M. P., Kovalenko, A. V. & Wallach, D. MAP3K-related kinase involved in NF- κ B induction by TNF, CD95 and IL-1. *Nature* **385**, 540–544 (1997).
- Ling, L., Cao, Z. & Goeddel, D. V. NF- κ B-inducing kinase activates IKK- α by phosphorylation of Ser-176. *Proc. Natl Acad. Sci. USA* **95**, 3792–3797 (1998).
- Guo, D. & Donner, D. B. Tumour necrosis factor promotes phosphorylation and binding of IRS-1 to phosphatidylinositol 3-kinase in 3T3-L1 adipocytes. *J. Biol. Chem.* **271**, 615–618 (1996).
- Alessi, D. R., Caudwell, F. B., Andjelkovic, M., Hemmings, B. A. & Cohen, P. Molecular basis for the substrate specificity of protein kinase B; comparison with MAPKAP kinase-1 and p60 S6 kinase. *FEBS Lett.* **399**, 333–338 (1996).
- Tanaka, M. *et al.* Embryonic lethality, liver degeneration, and impaired NF- κ B activation in IKK- β deficient mice. *Immunity* **10**, 421–429 (1999).
- Li, Q., Antwerp, D. V., Mercurio, F., Lee, K.-F. & Verma, I. M. Severe liver degeneration in mice lacking the I κ B kinase 2 gene. *Science* **284**, 321–326 (1999).
- Hu, Y. *et al.* Abnormal morphogenesis but intact IKK activation in mice lacking the IKK α of I κ B kinase. *Science* **284**, 316–320 (1999).
- Delhase, M., Hayakawa, M., Chen, Y. & Karin, M. Positive and negative regulation of I κ B kinase activity through IKK β subunit phosphorylation. *Science* **284**, 309–313 (1999).
- Ruderman, N. B., Kapeller, R., White, M. F. & Cantley, L. C. Activation of phosphatidylinositol 3-kinase by insulin. *Proc. Natl Acad. Sci. USA* **87**, 1411–1415 (1990).

Acknowledgements

L.D.M. is supported by a Hematology Oncology Training Grant from NIH. This work was supported by grants from NIH to D.B.D. and L.M.P.

Correspondence and requests for materials should be addressed to D.B.D. (e-mail: ddonner@IUPUI.edu).

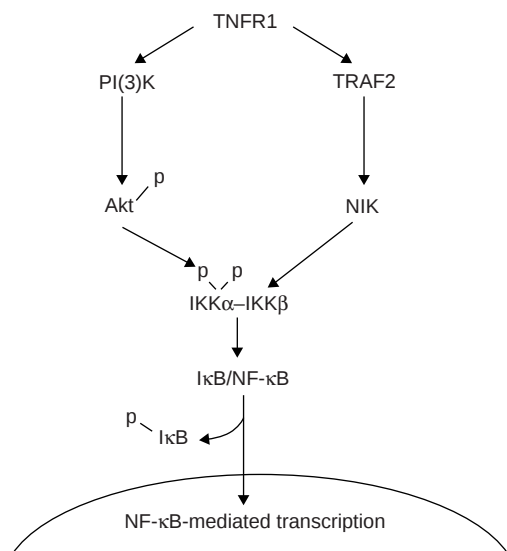


Figure 6 TNFR1 induces PI(3)K–Akt and TRAF2–NIK signalling pathways that activate IKK α at threonine 23 and serine 176, respectively.

NF- κ B is a target of AKT in anti-apoptotic PDGF signalling

Julia A. Romashkova & Sergei S. Makarov

Thurston Arthritis Research Center, and Center for Inflammatory Disorders, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7280, USA

The mechanisms of cell proliferation and transformation are intrinsically linked to the process of apoptosis: the default of proliferating cells is to die unless specific survival signals are provided^{1,2}. Platelet-derived growth factor (PDGF) is a principal survival factor that inhibits apoptosis and promotes proliferation³, but the mechanisms mediating its anti-apoptotic properties are not completely understood. Here we show that the transcription factor NF- κ B³⁻⁵ is important in PDGF signalling. NF- κ B transmits two signals: one is required for the induction of proto-oncogene *c-myc* and proliferation, and the second, an anti-apoptotic signal, counterbalances *c-Myc* cytotoxicity. We have traced a putative pathway whereby PDGF activates NF- κ B through Ras and phosphatidylinositol-3-kinase (PI(3)K) to the PKB/Akt protein kinase and the I κ B kinase (IKK); NF- κ B thus appears to be a target of the anti-apoptotic Ras/PI(3)K/Akt pathway^{6,7}. We show that, upon PDGF stimulation, Akt transiently associates *in vivo* with IKK and induces IKK activation. These findings establish a role for NF- κ B in growth factor signalling and define an anti-apoptotic Ras/PI(3)K/Akt/IKK/NF- κ B pathway, thus linking anti-apoptotic signalling with transcription machinery.

c-Myc is a central regulator of cell growth, death and differentiation. *c-Myc* is required for cell proliferation but, in the absence of survival factors, it induces apoptosis^{1,8}. PDGF inhibits *c-Myc*-induced apoptosis by activating the Ras/PI(3)K/Akt^{6,7} pathway, but the downstream effectors of this pathway are poorly defined. One consequence of PDGF treatment is the activation of NF- κ B⁹. NF- κ B controls the expression of numerous genes involved in inflammation, tumour development and immune responses³⁻⁵. In unstimulated cells, NF- κ B is retained in the cytoplasm through an interaction with inhibitory proteins known as I κ B. The canonical mechanism of NF- κ B activation involves signal-inducible degradation of I κ B, which causes the translocation of NF- κ B to the nucleus and transcription activation of targeted genes. As NF- κ B has been implicated in the regulation of apoptosis⁴, proliferation¹⁰ and *c-Myc* expression¹¹, we investigated a role for NF- κ B in PDGF signalling.

We used two types of primary fibroblast: normal human skin fibroblasts (NHF) and rat fibroblast-like synoviocytes (FLS). PDGF treatment induced NF- κ B DNA-binding activity, which became apparent after 1 h of stimulation, peaked at 3 h (Fig. 1a) and declined thereafter (data not shown). Supershift analysis has shown that the induced NF- κ B was a RelA/p50 heterodimer (data not shown). The induction of NF- κ B DNA-binding activity coincided with the degradation of I κ B α (Fig. 1a). To examine the role of NF- κ B in mitogenic activity of PDGF, NF- κ B activation was arrested by infecting the cells with an adenoviral vector encoding a non-degradable ('super-repressor') form of I κ B α (AdsrI κ B α)¹² (Fig. 1a). This inhibited DNA synthesis (Fig. 1b) and abrogated the induction of *c-Myc* message (data not shown) and protein (Fig. 1c). Thus, NF- κ B activation is required for the expression of *c-Myc* and proliferation.

To investigate the function of NF- κ B in anti-apoptotic PDGF signalling, exogenous *c-Myc* and super-repressor I κ B α (srI κ B α) were constitutively expressed in serum-deprived cells. To facilitate

detection, a reporter vector encoding green fluorescent protein (EGFP) was co-transfected with the *c-Myc* and srI κ B α vectors. In the absence of PDGF, constitutive *c-Myc* expression caused widespread death (Fig. 2a); the transfected cells had a characteristic apoptotic appearance, being round and condensed (data not shown). As expected, PDGF rescued the cells from *c-Myc*-induced apoptosis, but the co-expression of srI κ B α abrogated the rescue (Fig. 2a). To minimize the influence of variability in transfections and the observation error, we modified the apoptotic assay. In a modified experiment, we divided the transfected cells between two dishes and investigated the time course of cell viability. *c-Myc*-induced cell death was delayed and inhibited by PDGF (Fig. 2b, left), but the expression of srI κ B α abrogated this protection (Fig. 2b, right). In the absence of *c-Myc*, the expression of srI κ B α did not affect cell viability regardless of the presence of PDGF (Fig. 2b, right). These data indicate that NF- κ B activation is required for the anti-apoptotic activity of PDGF. To examine the anti-apoptotic effects of NF- κ B activation, exogenous NF- κ B (p50 and RelA subunits) was co-expressed with *c-Myc* in serum-deprived cells. Although p50 alone had no apparent effect, the expression of RelA alone or in a combination with p50 increased cell viability by 3 to 5-fold (Fig. 2c), indicating that NF- κ B activation itself can inhibit *c-Myc*-induced apoptosis.

How does PDGF activate NF- κ B? As the Ras/PI(3)K/Akt pathway is critical in anti-apoptotic PDGF signalling, we examined the possible role of this pathway. The expression of a dominant negative (DN) RasN17 mutant strongly inhibited PDGF-induced NF- κ B activation (Fig. 3a), indicating that Ras may be involved in NF- κ B activation. In anti-apoptotic PDGF signalling, PI(3)K is the proximal downstream target of Ras¹³. Wortmannin and LY294002, specific inhibitors of PI(3)K, both inhibited NF- κ B activation (Fig. 3b), indicating that PI(3)K may be involved in PDGF-induced NF- κ B activation. Importantly, neither DN Ras nor wortmannin

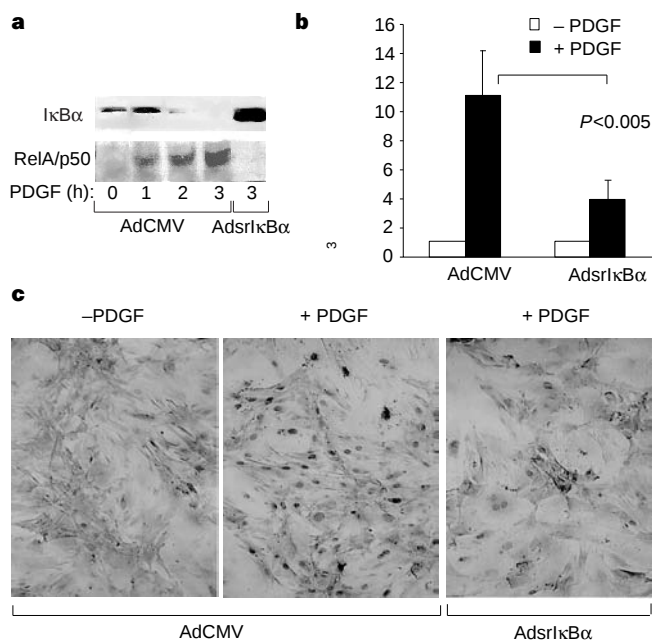


Figure 1 The mitogenic function of NF- κ B. FLS were infected with Ad vectors as indicated, incubated for 48 h in a serum-free medium and stimulated with PDGF. **a**, Time course of NF- κ B activation. I κ B α protein (top) and NF- κ B DNA-binding activity (bottom) were determined in the cytosolic and nuclear extracts. **b**, NF- κ B activation is required for DNA synthesis. [³H]Thymidine incorporation was determined after 48 h PDGF stimulation. Mean values $\pm$ s.e.m. are shown from three experiments performed in duplicate. Identical results were obtained in NHF (data not shown). **c**, NF- κ B activation is required for *c-Myc* protein expression. *c-Myc* immunostaining was performed after 24 h PDGF stimulation. Original magnification 200 $\times$.

affected the induction of NF- κ B by tumour necrosis factor- α (TNF α) and 12-*O*-tetradecanoylphorbol-13-acetate (TPA), stimuli that probably do not involve Ras and PI(3)K (Fig. 3c).

As Akt is a proximal downstream effector of the anti-apoptotic Ras/PI(3)K pathway^{6,7,13}, we next examined the involvement of Akt in NF- κ B activation. To this end, we used a kinase-dead AktK179M mutant¹⁴, which effectively inhibited the activation of Akt by PDGF (data not shown). The expression of AktK179M strongly inhibited NF- κ B reporter-gene expression, indicating that Akt is involved in NF- κ B activation (Fig. 3d). Similarly to the RasN17 and PI(3)K inhibitors, AktK179M expression did not affect cell viability (data not shown), thus ruling out the possibility that the inhibition of NF- κ B reporter-gene expression reflected a cytotoxic effect. Our data indicate that, in PDGF signalling, NF- κ B is a target of the Ras/PI(3)K/Akt pathway.

We next investigated the link between Akt and NF- κ B. PDGF treatment causes degradation of I κ B α (Fig. 1a), which indicates that IKK is involved. The inducible degradation of I κ B α occurs through several consecutive events, an initial step being the phosphorylation of I κ B α by a multisubunit IKK complex, the signalsome⁵. Two catalytic subunits of IKK, IKK α and IKK β , phosphorylate I κ B α on residues Ser 32 and Ser 36 and target I κ B α to ubiquitination and proteasomal degradation^{5,15}. To inhibit IKK, we used DN mutants of the IKK α and IKK β subunits. Whereas DN IKK α exerted only minimal (if any) inhibitory effect (data not shown), the expression of DN IKK β strongly inhibited PDGF-induced NF- κ B reporter-

gene expression (Fig. 3e). This prompted us to explore the causal relationship between Akt and IKK in NF- κ B activation. We found that the expression of a constitutively active myristylated Akt (myrAkt) alone was sufficient for NF- κ B activation (12-fold induction in FLS (Fig. 3f) and 7-fold induction in NHF (data not shown)). The induction of NF- κ B reporter by myrAkt was consistently observed in other cell types, including NIH 3T3 fibroblasts (data not shown). DN IKK β attenuated myrAkt-induced NF- κ B activation (Fig. 3f), indicating that IKK is a target of Akt in PDGF signalling. To investigate the link between Akt and IKK, we assessed the temporal relationship between these kinases. To assess the kinetics of Akt activation, we determined its phosphorylation on residues Thr 308 and Ser 473, which is a prerequisite for the catalytic activity of Akt¹³. Akt phosphorylation was apparent after 5 min of stimulation, reached a peak at 60 min and declined thereafter (Fig. 4a). These kinetics overlapped with the time course of IKK activation: the phosphorylation of I κ B α on Ser 32 was evident 30 min after PDGF stimulation and persisted for at least 2 h (Fig. 4b and data not shown). These data indicate that IKK might

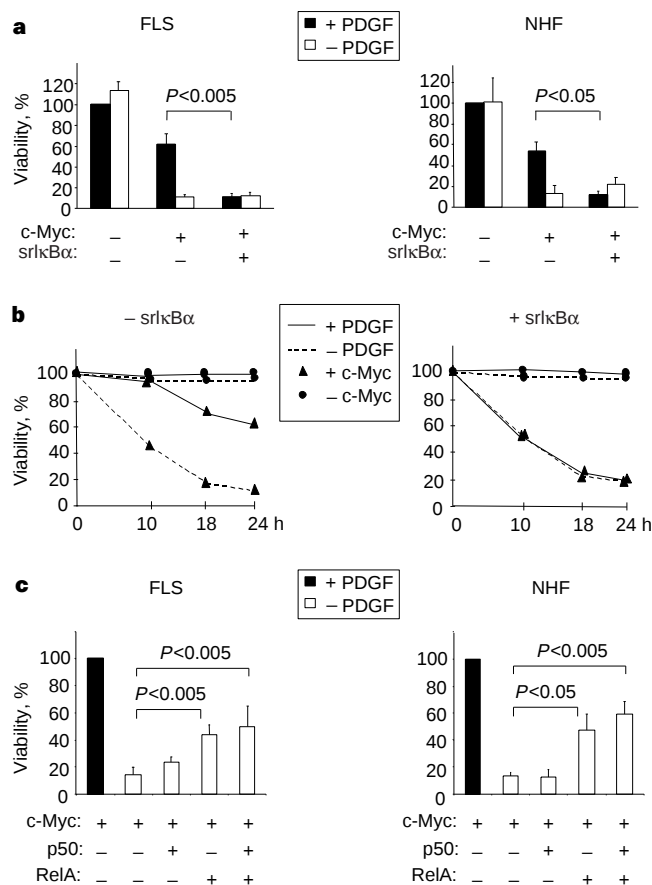


Figure 2 The anti-apoptotic function of NF- κ B. **a**, NF- κ B activation is required for the anti-apoptotic activity of PDGF. Cell viability was assessed after 18 h incubation in a low-serum medium. Bars represent mean values $\pm$ s.e.m. from four experiments in FLS (left) and one experiment in NHF (right). **b**, Time course of c-Myc-induced apoptosis in FLS in the absence (left) or presence (right) of srlkB α . **c**, The overexpression of NF- κ B inhibits c-Myc-induced apoptosis. Cell viability was assessed after 18-h incubation in a low serum medium. Bars represent mean values $\pm$ s.e.m. from two experiments in each of FLS (left) and NHF (right).

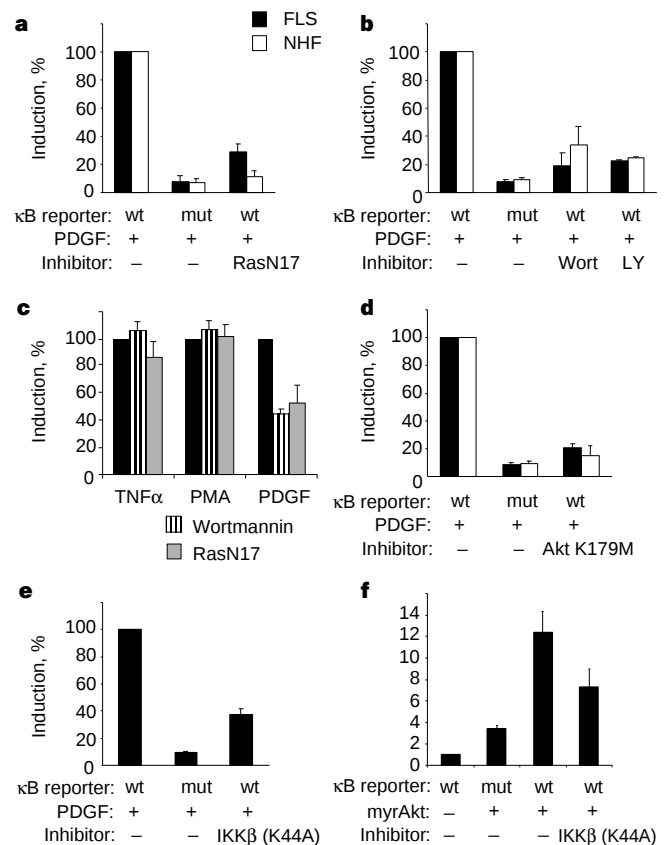


Figure 3 NF- κ B is a target of a Ras/PI(3)K/Akt/IKK pathway. Unless otherwise noted, transfected cells (5 μ g of each vector) were serum-starved for 48 h and stimulated with PDGF for 18 h. Bars represent mean values $\pm$ s.e.m. from *n* individual experiments performed in duplicate. **a**, The involvement of Ras. PDGF-induced activation of 3 κ BCAT in individual experiments was in the range of 8- to 37-fold in FLS (*n* = 4) and 9- to 15-fold in NHF (*n* = 3). **b**, The involvement of PI(3)K. Wortmannin (100 nM) and LY924002 (10 μ M) were added 30 min before PDGF (FLS, *n* = 2; NHF, *n* = 1). **c**, Inhibitors of Ras/PI(3)K do not affect TNF α - and PMA-induced NF- κ B activation. FLS were stimulated with TNF α (50 ng ml⁻¹), PMA (50 ng ml⁻¹) or PDGF for 6 h. Stippled bars: DN Ras-transfected cells. Hatched bars: wortmannin-pretreated cells. TNF α -, PMA- and PDGF-induced wild-type 3 κ BCAT activation was 37-, 11- and 8-fold, respectively; these values were accepted for 100% induction (solid bars) (FLS, *n* = 2). **d**, The involvement of Akt (FLS, *n* = 2; NHF, *n* = 2). **e**, The involvement of IKK (FLS, *n* = 3). **f**, NF- κ B activation by myrAkt. FLS were transfected with 2.5 μ g of each vector. Mean values $\pm$ s.d. from a single experiment performed in duplicate are shown. Identical data were obtained in NHF (not shown).

be a direct target of Akt, an assumption reinforced by the observation that PDGF-mediated phosphorylation of I κ B α did not require *de novo* protein synthesis (Fig. 4b and data not shown).

To assess the possible interaction between Akt and the IKK signalsome *in vivo*, we immunoprecipitated endogenous Akt and analysed the immune complexes for the presence of IKK using a combination of Akt and IKK α antibodies. The immunodetection revealed two bands corresponding to Akt (relative molecular mass (M_r) = 63K) and IKK α (M_r = 85K; Fig. 4c). These bands were absent when the immune complexes were precipitated by preadsorbed Akt antibodies, indicating the specificity of detection (Fig. 4c). Interestingly, the association of IKK and Akt was detected only in cells stimulated with PDGF for 1 h, not in unstimulated cells or those stimulated for 4 h. Consistent with the data shown in Fig. 4a, we found that the immunoprecipitated Akt was phosphorylated at 1 h of stimulation, and, to a much lesser degree, at 4 h (Fig. 4c). Similar kinetics of Akt and IKK association were observed in NIH 3T3 fibroblasts (Fig. 4d). In these experiments, we detected IKK in Akt immune complexes using an antibody against IKK β , another catalytic subunit of the IKK complex. Furthermore, the immunoprecipitation of endogenous IKK β in PDGF-stimulated NIH 3T3 cells brought down endogenous Akt (Fig. 4d). Combined, these observations show that, upon PDGF stimulation, Akt and IKK transiently associate *in vivo*.

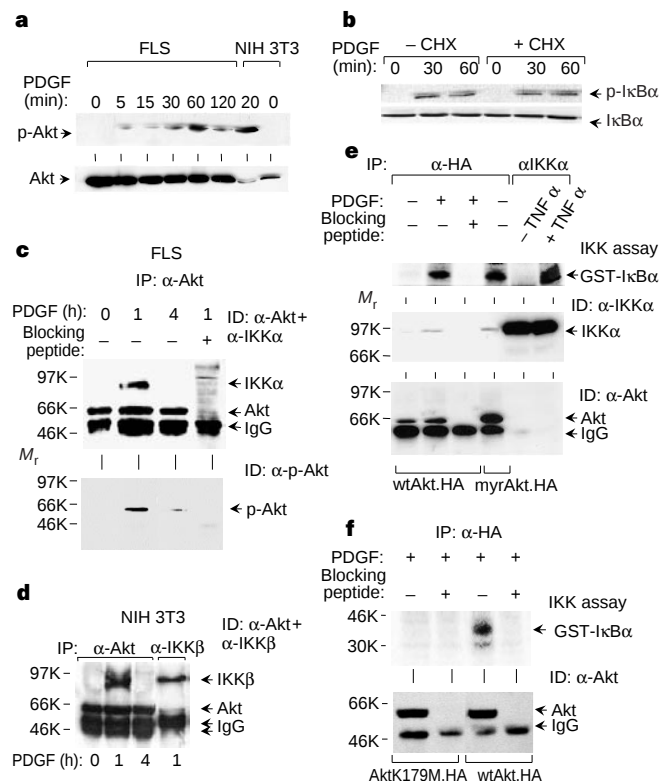


Figure 4 Akt associates with IKK *in vivo* and induces IKK activation. **a**, Akt activation. FLS lysates were consecutively immunodetected with phospho-Akt (Ser 473) (top) and total Akt (bottom) antibodies. NIH 3T3 lysates were from the antibody manufacturer. **b**, I κ B α phosphorylation. FLS lysates were consecutively immunodetected with p-I κ B α (Ser 32) (top) and total I κ B α (bottom) antibodies. CHX, cycloheximide. **c**, The association of Akt and IKK in FLS. Akt immune complexes were consecutively immunodetected with primary Akt and IKK α (top) and p-Akt (bottom) antibodies. M_r , relative molecular mass. **d**, The association of Akt and IKK in NIH 3T3 cells. Akt and IKK β immune complexes were immunodetected with primary Akt and IKK β antibodies. **e**, Akt activates IKK. HA-Akt and IKK α immune complexes were assayed for IKK activity (top). The 35K GST-I κ B α band is shown. The blot was consecutively immunodetected with IKK α (middle) and Akt (bottom) antibodies. **f**, Catalytic activity of Akt is required for IKK activation. IKK activity and Akt protein were assessed as in **e**.

To determine whether this association was functional (induced the catalytic activity of IKK) NIH 3T3 cells were transfected with haemagglutinin (HA)-tagged Akt vectors, stimulated with PDGF for 1 h, immunoprecipitated with anti-HA antibodies and assessed for I κ B kinase activity. As a positive control, endogenous IKK activity was determined in TNF α -stimulated untransfected cells. In cells transfected with wild-type (wt)Akt.HA, IKK activity was detected only in HA immune complexes from PDGF-treated cells, whereas myrAkt.HA-transfected cells contained IKK activity without stimulation (Fig. 4e). Conversely, IKK activity was undetectable in PDGF-stimulated cells transfected with the kinase-dead AktK179M.HA (Fig. 4f). The IKK activity correlated with the presence of IKK protein (Fig. 4e). Notably, Akt was undetectable in IKK immune complexes from TNF α -stimulated cells (Fig. 4e). This indicates that, in contrast to PDGF, TNF α does not induce the association of Akt with IKK. Together, our data indicate that PDGF-stimulated activation of NF- κ B is mediated by transient association of Akt with IKK and the activation of IKK.

Our study indicates that, in PDGF signalling, proliferative and pro-apoptotic (c-Myc induction) and anti-apoptotic signals bifurcate at the level of NF- κ B (Fig. 5). Consistent with the 'dual-signal' hypothesis postulating coupling of the proliferation pathways with those of cell death^{1,2}, NF- κ B has a dual role: it mediates the induction of pro-apoptotic c-Myc, and delivers an anti-apoptotic signal counterbalancing c-Myc cytotoxicity, conceivably by inducing the expression of a protective gene (or multiple genes). Given the roles of *ras*, *myc* and *akt* in cancer, our study has evident ramifications for understanding the mechanisms of cell transformation. It may explain why the ability of Ras to transform cells depends on NF- κ B activation¹⁶. Oncogenic Ras activates two independent executive mechanisms: an anti-apoptotic PI(3)K/Akt pathway and a pro-apoptotic Raf/MAPK pathway^{6,17}. Our results indicate that inhibition of NF- κ B might block anti-apoptotic Akt signalling without interfering with the pro-apoptotic pathway. Our observations also indicate a novel role for NF- κ B in rheumatoid arthritis. One feature of this disease is aggressive hyperplasia of the synovium associated with overexpression of c-Myc and activation of NF- κ B¹⁸. As PDGF is a principal growth factor in rheumatoid arthritis FLS cells, NF- κ B activation may contribute to synovial hyperplasia by promoting proliferation and inhibiting c-Myc-induced apoptosis. Our data identify NF- κ B as target of the anti-apoptotic Ras/PI(3)K/Akt pathway. Several putative targets of this pathway have been proposed, among them Bad^{19,20}, caspase-9²¹ and the transcription factor Forkhead^{22,23}. As all these targets represent proteins with intrinsic pro-apoptotic properties, Akt appeared to be a negative regulator of the cell-death machinery. Our data indicate that Akt can also promote survival by activating anti-apoptotic NF- κ B signalling. The interplay of these mechanisms in the anti-apoptotic activities of Akt¹³ has yet to be determined. The mechanism whereby PDGF activates NF- κ B also warrants further investigation. Our observations indicate that this mechanism involves consecutive activation of Akt and IKK, which is consistent with the observed phosphorylation and degradation of I κ B α and induction of NF- κ B DNA-binding activity. However, under certain circumstances, PI(3)K can activate NF- κ B through a mechanism that does not involve I κ B degradation²⁴. There is also evidence that Akt can activate NF- κ B transcriptional function through an alternative I κ B-independent

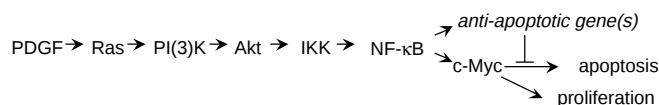


Figure 5 NF- κ B in PDGF signalling. PDGF activates the Ras/PI(3)K/Akt pathway. Activated Akt transiently associates with IKK and induces the activation of IKK and NF- κ B. NF- κ B transmits a signal required for the induction of c-Myc expression and proliferation, and an anti-apoptotic signal that inhibits c-Myc cytotoxicity.

mechanism, by modulating the transcriptional potential of RelA (M. Mayo, L. Madrid and A. Baldwin, personal communication). The contribution of these pathways remains to be clarified. Another issue is the role for the atypical isoforms of protein kinase C (PKC) and reactive oxygen intermediates (ROI), both of which have been implicated in mitogenic signalling^{25,26} and NF- κ B activation^{3,4}. As the atypical forms of PKC can associate with both Akt²⁷ and IKK²⁸, this indicates possible cooperation between PKC and Akt. Together, our findings indicate that the association of Akt and IKK may represent a distinct channel for NF- κ B activation, similar to those provided by NIK and MEKK1 in TNF α and IL-1 signalling^{5,29}. □

Methods

Reagents.

Unless otherwise noted, human recombinant PDGF-BB (Gibco BRL) was used at a final concentration of 50 ng ml⁻¹. Plasmids wtAkt.HA, AktK179M.HA and myrAkt.HA were gifts from P. Tsichlis; RasN17 was a gift from C. Der; and IKK α (K44A) and IKK β (K44A) were gifts from J. Woronicz. The CMV RelA, CMV p50, CMV srIkB α , CMV c-Myc and 3 κ BCAT vectors were gifts from A. Baldwin. The EGFP vector was from Promega. All plasmids were purified from bacterial endotoxin using Detoxigel (Pierce). Adenoviral infections were performed as described¹².

Cell culture.

We obtained primary FLS from rats with streptococcal cell wall (SCW)-induced arthritis³⁰. The cells were maintained in RPMI 1640/15% FBS and used at passages 5–18. Primary foreskin NHF were a gift from W. Kaufmann. The cells were maintained in RPMI 1640/10% FBS and used at passages 12–20.

Proliferation assay.

Ad-infected FLS (10⁴ cells per well) were starved for 48 h and stimulated with PDGF for another 48 h. [³H]thymidine (0.5 μ Ci per well) was added during the last 24 h. In individual experiments, incorporated radioactivity in PDGF-stimulated AdCMV-infected cells was in the range of 3,000 to 6,000 c.p.m.

NF- κ B assays.

We determined NF- κ B DNA-binding activity in FLS using electromobility-shift assay (EMSA) as described¹². The specificity of binding was determined by competition with a cold probe and by supershift analysis with anti-p50 and anti-RelA antibodies. To assess NF- κ B reporter-gene expression, cells (10⁵ cells per 100 mm dish) were transfected with wild-type or mutant 3 κ BCAT reporter vectors and corresponding DN inhibitors. Unless otherwise noted, each transfection contained 180 μ g of SuperFect reagent (Quiagen) and 5 μ g of each reporter and expression vector. The amount of DNA was kept constant by addition of an empty CMV plasmid. CAT activity was determined using the Qant-T-CAT kit (Amersham) and normalized on total protein. In some experiments, we used 3 κ BC LUC reporters; the luciferase assays were in good agreement with CAT assays.

Apoptotic assays.

Cells were co-transfected in a complete medium with a reporter EGFP along with c-Myc and srIkB α expression vectors (2 μ g of each DNA per 100 mm dish). Attached fluorescent cells were counted in three randomly chosen fields after 18 h incubation in a low serum (0.1% FBS) medium. In each experiment, at least 400 fluorescent cells were scored in PDGF-treated control (EGFP only) cells; these numbers were accepted for 100% viability. Alternatively, transfected FLS were divided between two dishes, allowed to adhere for 24 h and incubated in the low serum medium. Three randomly chosen observation fields were marked on the dishes, and the numbers of attached fluorescent cells were counted at indicated time points. To assess the influence of NF- κ B overexpression on c-Myc apoptosis, EGFP and c-Myc were co-transfected with p50 and RelA expression vectors (1.5 μ g of each DNA per 100 mm dish), and the viability was assessed after 18 h incubation in low serum medium. At least 250 fluorescent cells were scored in PDGF-treated control cells (EGFP plus c-Myc) (accepted for 100% viability).

Immunodetection.

For c-Myc immunostaining we used primary c-Myc antibodies (sc-764, Santa Cruz Biotechnology) (1:200 dilution) and a secondary HRP-conjugated antibody. For western blotting, lysates were resolved on SDS-PAGE and transferred to nitrocellulose. Total and phosphorylated IkB α were detected with primary IkB α antibodies (Rockland Inc.) (1:1000) and p-IkB α (Ser 32) antibodies (1:1000) (New England Biolabs). Where indicated, cells were pre-treated for 1 h with cycloheximide (CHX) (10 μ g ml⁻¹); to prevent degradation of phospho-IkB α , proteasomal inhibitor MG132 (5 μ M) was added with CHX. Total and phosphorylated Akt were detected with primary Akt (1:1000) and p-Akt (Ser 473) (1:1000) antibodies (New England Biolabs). IKK α was detected using either goat polyclonal antibodies against the carboxy-terminus of IKK α (1:1000) (sc-7120, Santa Cruz Biotechnology) or monoclonal antibodies against full-length IKK α (1.5 μ g ml⁻¹) (Pharmingen). IKK β was detected using antibodies against its C-terminus (1:400) (a gift from A. Manning). ECL detection was performed using an ECL kit (Du Pont NEN).

Immunoprecipitation.

Cells were lysed in a lysis buffer (5 mM HEPES pH 7.4, 150 mM NaCl, 1% Triton X-100, 10 mM glycerol, 1 mM EDTA, 2 mM Na₃VO₄, 5 mM phenylmethylsulphonyl fluoride (PMSF), 5 μ g ml⁻¹ aprotinin, 5 μ g ml⁻¹ leupeptin and 5 μ g ml⁻¹ pepstatin). Endogenous Akt was immunoprecipitated by 1-h incubation with an anti-Akt antibody (NEB) (1 μ g per 1 mg of protein) and 50 μ l of protein A Sepharose (Pharmacia). Where indicated, the antibody was preadsorbed with an Akt peptide (a gift from T. Davis). Endogenous IKK was immunoprecipitated by 1-h incubation with 1 μ g of an antibody against the C-terminus of IKK β and 50 μ l of protein A Sepharose. Immune complexes were washed five times with lysis buffer, boiled and subjected to SDS-PAGE.

IkB kinase assay.

NIH 3T3 cells were transiently transfected with HA-tagged Akt vectors, and starved by incubating in a 0.5% FBS medium for 48 h, followed by an overnight incubation in a serum-free medium. Cells were stimulated for 1 h with PDGF (100 ng ml⁻¹) and immunoprecipitated with HA antibodies (Y-11, Santa Cruz Biochemicals (SCB)) (1 μ g per 1 mg cell protein). Where indicated, the antibody was pre-adsorbed with an HA peptide (SCB). In control reactions, serum-starved untransfected cells were stimulated for 10 min with TNF α (100 ng ml⁻¹) and cell lysates were immunoprecipitated with IKK α monoclonal antibodies (Pharmingen) (2 μ g per 1 mg cell protein). *In vitro* IkB kinase activity was performed with recombinant GST-IkB α (1–54) as a substrate, as described¹⁵. Reaction products were separated by 10% SDS-PAGE, transferred to nitrocellulose and visualized with a phosphorimager. The blots were subsequently immunodetected for the presence of Akt and IKK proteins.

Statistics.

P values of comparisons were calculated using a paired two-tailed *t*-test.

Received 6 May; accepted 21 June 1999.

- Harrington, E. A., Bennett, M. R., Fanidi, A. & Evan, G. I. c-Myc-induced apoptosis in fibroblasts is inhibited by specific cytokines. *EMBO J.* **13**, 3286–3295 (1994).
- Evan, G. Cancer—a matter of life and cell death. *Int. J. Cancer* **71**, 709–711 (1997).
- Baldwin, A. S. Jr. The NF- κ B and I- κ B proteins: new discoveries and insights. *Annu. Rev. Immunol.* **14**, 649–683 (1996).
- Yin Foo, S. & Nolan, G. P. NF- κ B to the rescue: RELS, apoptosis and cellular transformation. *Trends Genet.* **15**, 229–235 (1999).
- Mercurio, F. & Manning, A. M. Multiple signals converging on NF- κ B. *Curr. Opin. Cell Biol.* **11**, 226–232 (1999).
- Kauffmann-Zeh, A. *et al.* Suppression of c-Myc-induced apoptosis by Ras signalling through PI(3)K and PKB. *Nature* **385**, 544–548 (1997).
- Kennedy, S. G. *et al.* The PI 3-kinase/Akt signaling pathway delivers an anti-apoptotic signal. *Genes Dev.* **11**, 701–713 (1997).
- Evan, G. I. *et al.* Induction of apoptosis in fibroblasts by c-myc protein. *Cell* **69**, 119–128 (1992).
- Olahaw, N. E., Kowalik, T. F., Huang, E. S. & Pledger, W. J. Induction of NF- κ B-like activity by platelet-derived growth factor in mouse fibroblasts. *Mol. Biol. Cell* **3**, 1131–1139 (1992).
- Hinz, M. *et al.* NF- κ B function in growth control: regulation of cyclin D1 expression an dG0/G1-to-S-phase transition. *Mol. Cell Biol.* **19**, 2690–2698 (1999).
- Kessler, D. J., Duyao, M. P., Spicer, D. B. & Sonenshein, G. E. NF- κ B-like factors mediate interleukin 1 induction of c-myc gene transcription in fibroblasts. *J. Exp. Med.* **176**, 787–792 (1992).
- Miagkov, A. V. *et al.* NF- κ B activation provides the potential link between inflammation and hyperplasia in the arthritic joint. *Proc. Natl Acad. Sci. USA* **95**, 13859–13864 (1998).
- Downward, J. Mechanisms and consequences of activation of protein kinase B/Akt. *Curr. Opin. Cell Biol.* **10**, 262–267 (1998).
- Franke, T. F. *et al.* The protein kinase encoded by the Akt proto-oncogene is a target of the PDGF-activated phosphatidylinositol 3-kinase. *Cell* **81**, 727–736 (1995).
- Mercurio, F. *et al.* IKK-1 and IKK-2: cytokine-activated IkB kinases essential for NF- κ B activation. *Science* **278**, 860–866 (1997).
- Mayo, M. W. *et al.* Requirement of NF- κ B activation to suppress p53-independent apoptosis induced by oncogenic Ras. *Science* **278**, 1812–1815 (1997).
- Marte, B. M., Rodriguez-Viciana, P., Wennstrom, S., Warne, P. H. & Downward, J. R-Ras can activate the phosphoinositide 3-kinase but not the MAP kinase arm of the Ras effector pathways. *Curr. Biol.* **7**, 63–70 (1997).
- Firestein, G. S. & Manning, A. M. Signal transduction and transcription factors in rheumatic disease. *Arthritis Rheum.* **42**, 609–621 (1999).
- Datta, S. R. *et al.* Akt phosphorylation of BAD couples survival signals to the cell-intrinsic death machinery. *Cell* **91**, 231–241 (1997).
- del Peso, L., Gonzalez-Garcia, M., Page, C., Herrera, R. & Nunez, G. Interleukin-3-induced phosphorylation of BAD through the protein kinase Akt. *Science* **278**, 687–689 (1997).
- Cardone, M. H. *et al.* Regulation of cell death protease caspase-9 by phosphorylation. *Science* **282**, 1318–1321 (1998).
- Brunet, A. *et al.* Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell* **96**, 857–868 (1999).
- Kops, G. J. P. L. *et al.* Direct control of the Forkhead transcription factor AFX by protein kinase B. *Nature* **398**, 630–634 (1999).
- Beraud, C., Henzel, W. J. & Baeuerle, P. A. Involvement of regulatory and catalytic subunits of phosphoinositide 3-kinase in NF- κ B activation. *Proc. Natl Acad. Sci. USA* **96**, 429–434 (1999).
- Berra, E. *et al.* Protein kinase C- ζ isoform is critical for mitogenic signal transduction. *Cell* **74**, 555–563 (1993).
- Sundaresan, M., Yu, Z.-X., Ferrans, V. J., Irani, K. & Finkel, T. Requirement for generation of H₂O₂ for platelet-derived growth factor signal transduction. *Science* **270**, 296–299 (1995).
- Konishi, H., Shinomura, T., Kuroda, S., Ono, Y. & Kikkawa, U. Molecular cloning of rat RAC protein kinase α and β and their association with protein kinase C ζ . *Biochem. Biophys. Res. Commun.* **205**, 817–825 (1994).

28. Lallena, M. J., Diaz-Meco, M. T., Bren, G., Paya, C. V. & Moscat, J. Activation of κ B kinase beta by protein kinase C isoforms. *Mol. Cell. Biol.* **19**, 2180–2188 (1999).
29. Karin, M. & Delhase, M. JNK or IKK, AP-1 or NF- κ B, which are the targets for MEK kinase 1 action? *Proc. Natl Acad. Sci. USA* **95**, 9067–9069 (1998).
30. Makarov, S. S. *et al.* Suppression of experimental arthritis by gene transfer of interleukin 1 receptor antagonist cDNA. *Proc. Natl Acad. Sci. USA* **93**, 402–406 (1996).

Acknowledgements

We thank P. Tschlis, J. Woronicz, A. Manning, F. Mercurio, C. Der, T. Davis, W. Kaufmann, A. Danilkovitch and A. Baldwin for sharing reagents; A. Baldwin, C. Der and P. Cohen for stimulating discussions; C. Bradham for assistance in kinase assays; J. Watson for editorial assistance; and members of the S.M. laboratory, A. Miagkov and D. Kovalenko for technical assistance. This work was supported by grants from the National Institutes of Health and the Arthritis Foundation.

Correspondence and requests for materials should be addressed to S.M. (e-mail: smak@med.unc.edu).

Global unfolding of a substrate protein by the Hsp100 chaperone ClpA

Eilika U. Weber-Ban*, Brian G. Reid*†, Andrew D. Miranker‡ & Arthur L. Horwich*†

* Department of Genetics and † Howard Hughes Medical Institute, Yale University School of Medicine, Boyer Center, 295 Congress Avenue, New Haven, Connecticut 06510, USA

‡ Department of Molecular Biophysics and Biochemistry, Yale University, 260 Whitney Avenue, New Haven, Connecticut 06520, USA

The bacterial protein ClpA, a member of the Hsp100 chaperone family, forms hexameric rings that bind to the free ends of the double-ring serine protease ClpP (refs 1, 2). ClpA directs the ATP-dependent degradation of substrate proteins bearing specific sequences^{3–5}, much as the 19S ATPase 'cap' of eukaryotic proteasomes functions in the degradation of ubiquitinated proteins^{6–8}. In isolation, ClpA and its relative ClpX can mediate the disassembly of oligomeric proteins^{9,10}; another similar eukaryotic protein, Hsp104, can dissociate low-order aggregates¹¹. ClpA has been proposed to destabilize protein structure, allowing passage of proteolysis substrates through a central channel into the ClpP proteolytic cylinder^{12–14}. Here we test the action of ClpA on a stable monomeric protein, the green fluorescent protein GFP, onto which has been added an 11-amino-acid carboxy-terminal recognition peptide, which is responsible for recruiting truncated proteins to ClpAP for degradation^{5,15}. Fluorescence studies both with and without a 'trap' version of the chaperonin GroEL, which binds non-native forms of GFP¹⁶, and hydrogen-exchange experiments directly demonstrate that ClpA can unfold stable, native proteins in the presence of ATP.

We first tested whether GFP, a very stable monomeric protein (it is resistant to denaturation by 6M urea, for example; Fig. 1a), could be degraded by the ClpAP machinery once recruited to it as a result of the attachment of an 11-residue ssrA peptide to the GFP terminus. Appendage of the peptide did not interfere with the fluorescent properties or stability of the GFP portion: the tagged protein (GFP11) gave the same fluorescence quantum yield and emission spectrum as non-tagged GFP (not shown) and was as stable to denaturant (Fig. 1a). When incubated with catalytic amounts of ClpA and ClpP in the presence of ATP, however, GFP11 was completely degraded, as evidenced both by loss of fluorescence (Fig. 1b) and by the disappearance of the protein on SDS–polyacrylamide gel electrophoresis (Fig. 1c). The kinetics of the two measurements were similar, preventing the resolution of an

unfolding action from degradation. In contrast with the tagged protein, non-tagged GFP gave no change in fluorescence or loss of protein when incubated under similar conditions with ClpA, ClpP and ATP (results not shown). We conclude that even the stable β -barrel of GFP can be destabilized and degraded by the ClpAP machinery when recruited to it as a result of the addition of the ssrA sequence.

To test whether we could detect unfolding of GFP11 as a step distinct from the degradation carried out by ClpAP, we incubated GFP11 with stoichiometric amounts of ClpA on its own. In the presence of ATP, but not of ATP- γ S, there was a small decrease (20%) in fluorescence intensity (Fig. 2a), indicating that some fraction of the input GFP11 molecules could be unfolded. The observed effect was small, however, compared with the effect of adding a stoichiometric amount of ClpA and ClpP, which caused a complete loss of fluorescence as a result of the total proteolytic degradation of the added GFP11 (Fig. 2a, and results not shown).

We reasoned that, if GFP11 molecules unfolded by ClpA were released from it and then refolded, establishing an equilibrium between unfolding and refolding, then preventing refolding should reveal the unfolding action of ClpA, producing a larger drop in fluorescence. To prevent refolding, we added a GroEL 'trap' mutant (D87K, in which an aspartate residue at position 87 is replaced by a

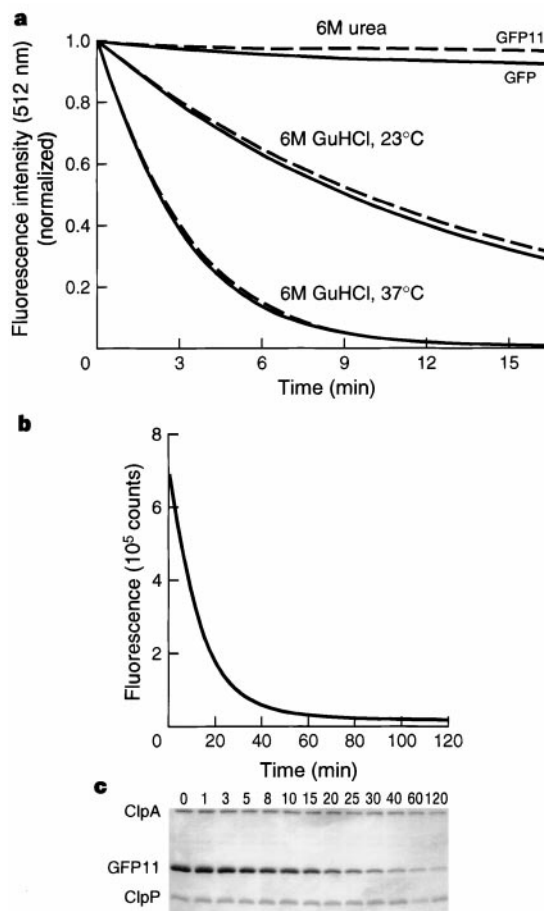


Figure 1 GFP11, a derivative bearing an 11-amino-acid C-terminal ssrA tag recognized by ClpA, exhibits the same stability to denaturants as untagged GFP, and is degraded by ClpA/ClpP in the presence of ATP. **a**, Time course of fluorescence intensity changes of GFP11 and GFP after dilution to 600 nM in 6 M urea at 23 °C (top), 6 M guanidine-HCl at 23 °C (middle), or 6 M guanidine-HCl at 37 °C (bottom). **b**, Time-course of the fluorescence intensity change of 10 μ M GFP11 incubated with 100 nM ClpA, 50 nM ClpP and 10 mM ATP. Excitation and emission wavelengths were 400 and 510 nm, respectively. **c**, SDS–PAGE analysis of aliquots of the reaction used in **b**, taken at the times indicated.

lysine) to the reaction: this mutant captures but is unable to release non-native forms of GFP. As shown in Fig. 2b, when a 5-fold molar excess of trap is present with ATP in a stoichiometric mixture of GFP11 and ClpA, there is a large and irreversible loss of fluorescence, unlike the small change in the absence of trap, indicating that efficient unfolding can be mediated by ClpA. Unfolding is not effected by the GroEL trap alone, which when incubated with GFP11 gave no fluorescence change (results not shown), nor does it occur in the presence of the non-hydrolysable ATP analogue ATP- γ S (Fig. 2c), which is consistent with the inability of this nucleotide to support turnover of GFP11 by ClpAP.

To confirm that ClpA was releasing non-native forms that were captured by the trap, we did the same experiment with 35 S-methionine-labelled GFP11, and subjected the mixture to gel filtration. In the absence of ClpA, but in the presence of trap and ATP, virtually all (92%) of the input molecules were found at the same migration position as GFP11 (Fig. 3a, b), which also indicates that the GroEL trap does not bind native GFP11. After addition of ClpA, however, most (80%) of the input molecules migrated at the position of trap (Fig. 3c), in agreement with the loss of fluorescence shown in Fig. 2b. The remaining molecules were found at the same

position as GFP11. They had either failed to be recognized by ClpA during the reaction or had dissociated from the trap during gel filtration. In control incubations, replacement of ATP with ATP- γ S or the omission of trap resulted in all the input labelled molecules migrating at the GFP11 position. This is consistent with ATP- γ S being unable to promote unfolding, with the refolding of GFP11 in the absence of trap, and with the unstable association of GFP11 with ClpA in gel-filtration chromatography. We found that GFP11 did not stably associate with ClpA even when chromatography was done in the presence of ATP (results not shown). We conclude that ClpA mediates ATP-dependent unfolding of GFP11, producing non-native forms that are devoid of fluorescence and that are readily recognized by the chaperonin GroEL. These non-native forms are probably collapsed globular states that expose hydrophobic surfaces which are recognizable by the GroEL trap apical domains.

To assess the competition between trap and ClpP for the ClpA-unfolded substrate protein, we supplied a stoichiometric amount of ClpP to ClpA with trap in 5-fold molar excess. The 35 S label did not elute with the trap in gel filtration but as a broad peak several minutes after GFP11 (result not shown), indicating that GFP11 had been degraded. In the presence of ClpP, therefore, GFP11 that has been unfolded by ClpA is probably released from ClpA into the adjoining ClpP proteolytic cavity, without access to trap, and is committed to degradation.

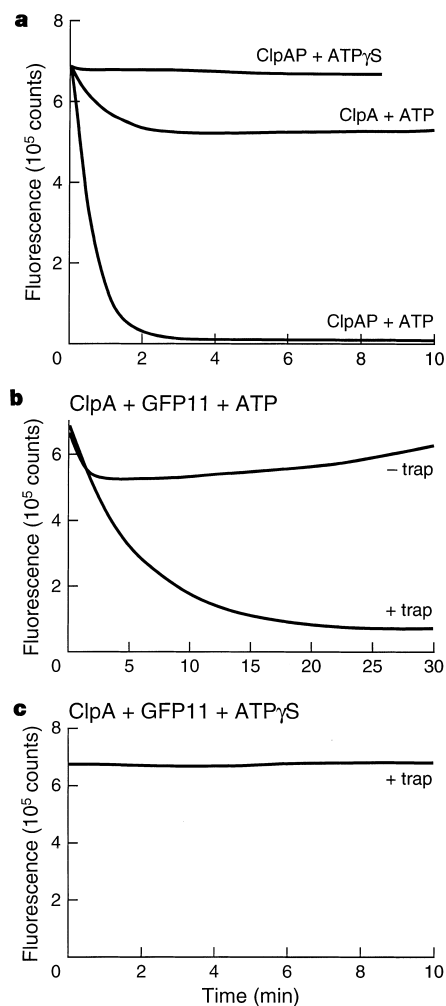


Figure 2 ClpA, in the absence of ClpP, mediates ATP-dependent unfolding of GFP11. **a**, Time course of fluorescence intensity changes of 2 μ M GFP11, incubated with 2 μ M ClpA and ATP; or with ClpA and ClpP, each 2 μ M, and ATP or ATP- γ S. **b**, Time course of fluorescence intensity changes of 2 μ M GFP11, in the presence of 2 μ M ClpA, ATP, and in the absence or presence of 10 μ M of the GroEL trap mutant D87K. The recovery of fluorescence intensity at later times in the absence of trap probably reflects exhaustion of the ATP-regeneration system. **c**, Lack of fluorescence intensity change of GFP11 in ATP- γ S even in the presence of trap, indicating that ATP- γ S does not support unfolding of GFP11.

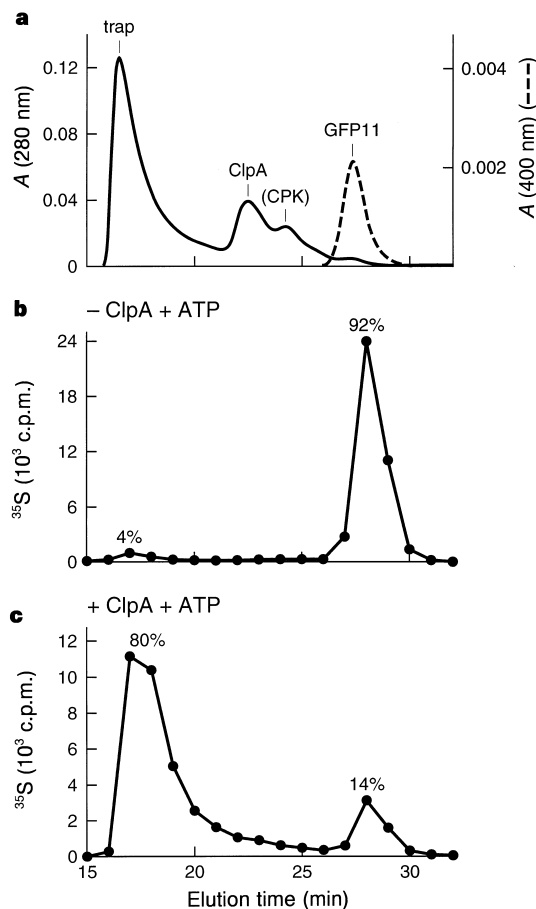


Figure 3 Transfer of GFP11 to GroEL trap in the presence of ClpA and ATP. **a**, Gel-filtration profile showing positions of migration of GroEL trap, ClpA (dimer) and CPK (from the ATP-regenerating system) (absorbance A at 280 nm) and GFP11 (A at 400 nm). **b**, Elution profile of 2 μ M 35 S-GFP11 after incubation for 40 min in the presence of ATP with a 5-fold molar excess of GroEL trap, showing that most of the GFP11 is uncomplexed. **c**, Elution profile of 2 μ M 35 S-GFP11 incubated with 2 μ M ClpA and ATP under the same conditions as in **b**, showing transfer of ~80% of the molecules to the trap. In all cases, chromatography was on a Superose-12 gel-filtration column (Pharmacia) and without ATP in the column buffer.

To determine whether the unfolding mediated by ClpA is partial, like the dissociation of a dimer, exposing hydrophobic surfaces at the interface but leaving the core intact, or total, perhaps destabilizing the core of a substrate protein, we used hydrogen–deuterium exchange and mass spectrometry experiments¹⁷. GFP11 was deuterated (dGFP11) by unfolding it in 6 M guanidine-DCl in D₂O and then refolded by diluting into D₂O buffer. Its molecular mass was determined after incubating with ATP in H₂O buffer (Fig. 4b) to be 106 daltons heavier than fully protonated GFP11 (27,861 daltons; Fig. 4a), reflecting the number of sites that are protected by the core in the folded state. Incubation with ClpA and ATP for 2 h at 23 °C, however, produced a molecular mass (Fig. 4c) that was within 6 daltons of fully protonated GFP11, indicating that nearly complete hydrogen exchange of the stable core of dGFP11 had occurred. By contrast, in a mixture of ClpA with ATP-γS (Fig. 4d), dGFP retained >100 daltons of deuterium. When dGFP11 was incubated with GroEL in the absence of nucleotide, the retention of deuterium was the same (results not shown), reflecting that the chaperonin neither stably binds (Fig. 3b) nor transiently unfolds native GFP11 on this timescale.

We conclude that ClpA in the presence of ATP mediates the unfolding of proteins bound after recognition of specific terminal amino-acid sequences. In association with ClpP protease, these destabilized structures are committed to translocation from the cavity of ClpA into the hydrolytic cavity of the protease, where they are degraded. In the absence of associated ClpP, substrate proteins are released from ClpA into solution, where they may spontaneously refold or become bound, as in the trap experiments described here, by other components such as chaperonins. This action of unfolding and release may also be exerted by other Hsp100 family members: for example, in the dissociation by ClpX of the

MuA transposase tetramer from DNA¹⁰, and in the dissociation by eukaryotic Hsp104 of low-order aggregates¹¹.

The geometry of unfolding and translocation by ClpA remains unclear. Presumably unfolding is driven by multivalent contacts between the surrounding ClpA subunits and the substrate protein. Although the nature of these contacts is unknown, ATP binding or hydrolysis probably drives a conformational change in ClpA that pulls the protein structure apart, in an action analogous to the transient unfolding of Rubisco after GroES and ATP bind to a Rubisco–GroEL binary complex¹⁸. ClpA releases the unfolded substrate, either into the bulk solution, as indicated by the trap experiment, or into the proteolytic cylinder of ClpP when it is present, whereas the unfolded protein released into the GroEL–GroES *cis* cavity is committed to a time-limited attempt at folding to the native state in this sequestered and hydrophilic environment. But how directional translocation through the ClpA cavity occurs during the unfolding action remains unknown. □

Methods

Proteins.

ClpA, ClpP, and GFP11 (Q80R) were all overproduced from T7 expression plasmids and purified by ion-exchange chromatography or as described (ClpP)¹⁹. GroEL D87K was expressed and purified as described²⁰. Concentrations refer to the fully assembled species.

Fluorescence.

Fluorescence kinetic measurements were made on a PTI Quantamaster QM-1 fluorimeter connected to a PC-based data acquisition program, with excitation at 400 nm and emission at 510 nm. In reactions with catalytic amounts of ClpA and ClpP (Fig. 1), reaction buffer was 50 mM HEPES, pH 7.5, 0.3 M NaCl, 20 mM MgCl₂, 0.5 mM DTT, 10% glycerol, plus 10 mM ATP. For all other reactions, where substrate and ClpA/ClpP were stoichiometric, the reaction buffer contained 0.05 U μl⁻¹ creatine phosphokinase, 30 mM phosphocreatine, 5 mM ATP. In all reactions, ClpA, with ClpP where indicated, was incubated with ATP and all other components for 20–30 s to enable ClpA and ClpAP complexes to assemble before addition of GFP11.

Hydrogen–deuterium exchange.

Deuterated GFP11 (dGFP11) was prepared by two cycles of exchange of protein into D₂O. In the first cycle, GFP11 was diluted 15-fold with D₂O and dried to its original concentration. In the second cycle, the protein was unfolded by mixing with 5 vol. 6 M guanidine-DCl (pD 7.5) and incubating at 55 °C for 2 min, then at 23 °C for 10 min. Complete unfolding was verified by total loss of fluorescence. The protein was refolded by 50-fold dilution into deuterated refolding buffer (25 mM Tris, pD 7.5, 2.5% glycerol, 4 mM DTT), followed by incubation for 2 h at 23 °C. Aggregates were removed by ultrafiltration (Centriprep 100). The filtrate was concentrated to ~100 μM by ultrafiltration (Microcon 10). All quoted pH measurements in D₂O buffers (pD) are uncorrected pH meter readings.

dGFP11 was diluted 50-fold into H₂O reaction mixtures, giving a final deuterium content of 2%. Reactions were carried out for 2 h at 23 °C. ClpA was then precipitated by 3-fold dilution with 7.5 mM Tris, pH 7.5, and removed by centrifugation. The supernatant was filtered through a Centricon 100 to remove residual ClpA, and the filtrate was concentrated to ~30 μM GFP11 with a Microcon 10. Buffer was exchanged into 7.5 mM Tris, pH 7.5, using a Microcon 10 before mass spectrometric analysis. These steps of sample preparation were carried out at 23 °C and required ~6 h.

Mass spectrometry.

Spectra were collected using a Micromass LCT electrospray time of flight mass spectrometer. Atmospheric pressure ionization was performed using borosilicate glass capillaries drawn and sputter coated in-house. All spectra show the +13 charge state expressed on a mass scale, but mass and Δ-mass determinations were made using a minimum of three charge states. All spectra were collected over a mass range of 200 to 5,000 *m/z*, signal-averaged for 1 min with no smoothing applied. 25–30 point external calibrations were done using 20 mM CsI in H₂O ionized under matched instrument conditions. GFP11 samples were kept on ice until analysis and diluted 10–50-fold with H₂O immediately before loading into the capillaries. The mass spectrometer ionization source temperature was kept at 20 °C.

Received 30 April; accepted 29 June 1999.

- Kessel, M. *et al.* Homology in structural organization between *E. coli* ClpA protease and the eukaryotic 26 S proteasome. *J. Mol. Biol.* **250**, 587–594 (1995).
- Beuron, F. *et al.* At sixes and sevens: characterization of the symmetry mismatch of the ClpAP chaperone-assisted protease. *J. Struct. Biol.* **123**, 248–259 (1998).
- Wickner, S. *et al.* A molecular chaperone, ClpA, functions like DnaK and DnaJ. *Proc. Natl Acad. Sci. USA* **91**, 12218–12222 (1994).
- Levchenko, I., Smith, C. K., Walsh, N. P., Sauer, R. T. & Baker, T. A. PDZ-like domains mediate binding specificity in the Clp/Hsp100 family of chaperones and protease regulatory subunits. *Cell* **91**, 939–947 (1997).

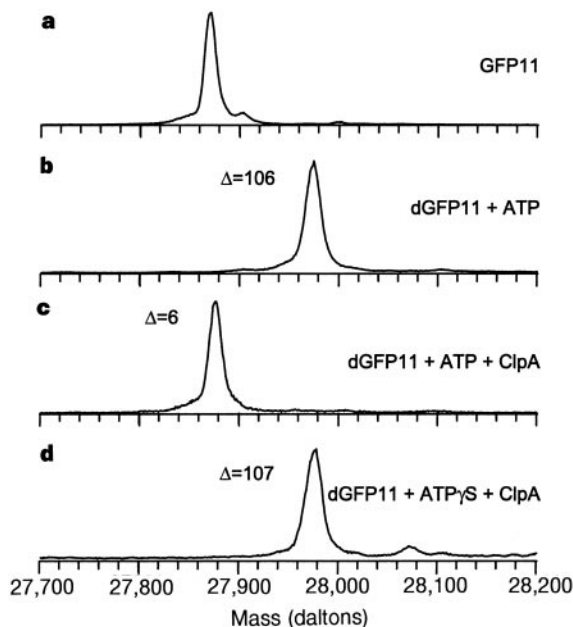


Figure 4 Mass spectra of GFP11 subjected to hydrogen–deuterium exchange followed by incubation of the deuterated protein in H₂O buffer with or without ClpA and nucleotide. **a**, Native GFP11, giving a mass of 27,861 daltons. This compares well with the calculated mass of 27,861 daltons. **b**, dGFP11 incubated in H₂O buffer with ATP for 2 h at 23 °C, giving a mass of 27,967 daltons. **c**, dGFP11 incubated as in **b**, but with ClpA and ATP, giving a mass of 27,867 daltons, amounting to a nearly complete loss of deuterium label. **d**, dGFP11 incubated as in **c**, but with ATP-γS instead of ATP, giving a mass of 27,968 daltons. In **b–d**, mass differences (Δ) are shown with respect to the mass measured in **a**. A representative experiment is shown, with all samples being measured on the same day. The mass changes shown here are within 6 daltons of the average difference in mass derived from three independent experiments.

5. Gottesman, S., Roche, E., Zhou, Y. N. & Sauer, R. T. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev.* **12**, 1338–1347 (1998).
6. DeMartino, G. N. *et al.* PA700, an ATP-dependent activator of the 20 S proteasome, is an ATPase containing multiple members of a nucleotide-binding protein family. *J. Biol. Chem.* **269**, 20878–20884 (1994).
7. Glickman, M. H. *et al.* A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and eIF3. *Cell* **94**, 615–623 (1998).
8. Baumeister, W., Walz, J., Zühl, F. & Seemüller, E. The proteasome: paradigm of a self-compartmentalizing protease. *Cell* **92**, 367–380 (1998).
9. Pak, M. & Wickner, S. Mechanism of protein remodeling by ClpA chaperone. *Proc. Natl Acad. Sci. USA* **94**, 4901–4906 (1997).
10. Levchenko, I., Luo, L. & Baker, T. A. Disassembly of the Mu transposase tetramer by the ClpX chaperone. *Genes Dev.* **9**, 2399–2408 (1995).
11. Glover, J. R. & Lindquist, S. Hsp104, Hsp70, and Hsp40: a novel chaperone system that rescues previously aggregated proteins. *Cell* **94**, 73–82 (1998).
12. Gottesman, S., Maurizi, M. R. & Wickner, S. Regulatory subunits of energy-dependent proteases. *Cell* **91**, 435–438 (1997).
13. Larsen, C. N. & Finley, D. Protein translocation channels in the proteasome and other proteases. *Cell* **91**, 431–434 (1997).
14. Wang, J., Hartling, J. A. & Flanagan, J. M. The structure of ClpP at 2.3 Å resolution suggests a model for ATP-dependent proteolysis. *Cell* **91**, 447–456 (1997).
15. Keiler, K. C., Waller, P. R. H. & Sauer, R. T. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* **271**, 990–993 (1996).
16. Weissman, J. S., Rye, H. S., Fenton, W. A., Beechem, J. M. & Horwich, A. L. Characterization of the active intermediate of a GroEL-GroES-mediated protein folding reaction. *Cell* **84**, 481–490 (1996).
17. Miranker, A., Robinson, C. V., Radford, S. E. & Dobson, C. M. Investigation of protein folding by mass spectrometry. *FASEB J.* **10**, 93–101 (1996).
18. Shtilerman, M., Lorimer, G. H. & Englander, S. W. Chaperonin function: folding by forced unfolding. *Science* **284**, 822–825 (1999).
19. Maurizi, M. R., Clark, W. P., Kim, S.-H. & Gottesman, S. ClpP represents a unique family of serine proteases. *J. Biol. Chem.* **265**, 12546–12552 (1990).
20. Fenton, W. A., Kashi, Y., Furtak, K. & Horwich, A. L. Residues in chaperonin GroEL required for polypeptide binding and release. *Nature* **371**, 614–619 (1994).

Acknowledgements

E.U.W. is a postdoctoral associate of the Jane Coffin Childs Foundation. A.D.M. is a Pew Scholar in the biomedical sciences. This work was supported by the NIH, the Howard Hughes Medical Institute and the Nelson Fund.

Correspondence and requests for materials should be addressed to A.L.H. (e-mail: horwich@csbmet.csb.yale.edu).

Structure of *Tetrahymena* GCN5 bound to coenzyme A and a histone H3 peptide

Jeannie R. Rojas^{*†}, Raymond C. Trievel^{*‡}, Jianxin Zhou[§], Yi Mo^{*†}, Xinmin Li^{*‡}, Shelley L. Berger^{*}, C. David Allis^{§||} & Ronen Marmorstein^{*†‡}

^{*} The Wistar Institute, [†] Department of Chemistry, and [‡] Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

[§] Department of Biology, University of Rochester, Rochester, New York 14627, USA

Gene activation is a highly regulated process that requires the coordinated action of proteins to relieve chromatin repression and to promote transcriptional activation. Nuclear histone acetyltransferase (HAT) enzymes provide a mechanistic link between chromatin destabilization and gene activation by acetylating the ϵ -amino group of specific lysine residues within the amino-terminal tails of core histones to facilitate access to DNA by transcriptional activators^{1,2}. Here we report the high-resolution crystal structure of the HAT domain of *Tetrahymena* GCN5 (tGCN5) bound with both its physiologically relevant ligands, coenzyme A (CoA) and a histone H3 peptide, and the structures of

nascent tGCN5 and a tGCN5/acetyl-CoA complex. Our structural data reveal histone-binding specificity for a random-coil structure containing a G-K-X-P recognition sequence, and show that CoA is essential for reorienting the enzyme for histone binding. Catalysis appears to involve water-mediated proton extraction from the substrate lysine by a glutamic acid general base and a backbone amide that stabilizes the transition-state reaction intermediate. Comparison with related N-acetyltransferases indicates a conserved structural framework for CoA binding and catalysis, and structural variability in regions associated with substrate-specific binding.

The structure of the nascent tGCN5 HAT domain was determined by multiple anomalous dispersion phasing, and the structures of the binary tGCN5/acetyl-CoA and ternary tGCN5/CoA/histone H3 peptide complexes were determined by molecular replacement using the refined tGCN5 model (Table 1). The tGCN5 HAT domain contains a mixed α/β structure with a globular fold and an overall structure that is very similar to its yeast (yGCN5)³ and human (hP/CAF)⁴ homologues (Fig. 1a). The central core region has sequence and structural homology to a group of over 50 N-acetyltransferases called GNAT (GCN5-related N-acetyltransferases)⁵, and the amino- and carboxy-terminal domains share sequence and structural divergence with several GNAT members^{6–9} (Fig. 1a). Between the N- and C-terminal segments of the protein are two pronounced and roughly orthogonal surface clefts forming an L-shape (Fig. 1b). In the binary tGCN5/acetyl-CoA complex, this cleft is occupied by acetyl-CoA, whereas in the ternary tGCN5/CoA/histone H3 peptide complex, the larger cleft is also occupied by the histone H3 peptide.

The ternary complex contains an 11-residue peptide from yeast histone H3 centred around the reactive Lys 14 and containing the sequence K₉-S₁₀-T₁₁-G₁₂-G₁₃-K₁₄-A₁₅-P₁₆-R₁₇-K₁₈-Q₁₉. The side chains for residues 9, 10 and 11 are not visible in the final electron-density map and are modelled as Ala₉-Ala₁₀-Gly₁₁ (Fig. 2a). The 11-residue histone H3 peptide is bound above the β 4 strand of the core domain and is flanked on opposite sides by the α 1-turn- α 2 N-terminal segment and the β 5-turn- α 4 and α 5-turn- β 6 C-terminal segments (Fig. 1a). The peptide adopts a random-coil structure and buries a total of 1,562 Å² of solvent-exposed surface area upon complex formation.

Most of the protein-peptide interactions are mediated through the backbone of the histone H3 peptide (Fig. 2b), and about 75% of these interactions involve Lys 14 and the five residues C-terminal to it. Protein-peptide interactions are distributed roughly evenly between hydrogen bond and van der Waals contacts, with most of the peptide interactions mediated by residues 159–165 (excluding residue 161) of the β 5-turn- α 4 region in the C-terminal protein segment. Residues 77–80 within the α 1-turn N-terminal segment also make peptide interactions, and all of these interactions (except for Pro 78) are with the backbone of the peptide. Tyr 192 of the α 5- β 6 loop C-terminal segment also contacts the peptide backbone, and CoA directly contacts Pro 16 of the peptide. The reactive Lys 14 residue of the histone H3 peptide is directly anchored to the substrate binding site by van der Waals interactions with Val 123 and Leu 126 in the β 4 strand at the base of the peptide cleft and by the CoA cofactor, and by N ζ hydrogen bonds to the backbone carbonyl group of Ala 124 and the terminal sulphur atom of CoA (Fig. 2b). Significantly, the region of the β 4 strand that interacts with the Lys 14 residue of the histone H3 peptide has a pronounced kink (Fig. 1a) which appears to be essential to facilitate these protein-substrate interactions as well as to point the backbone carbonyls of residues Val 123 and Ala 124 into the histone-binding cleft to facilitate an electrostatic attraction for the lysine substrate. The β 4 kink is found in each of the structures presented here, as well as in the related yGCN5 (ref. 3) and hP/CAF/CoA (ref. 4) structures and in the analogous strands of the functionally divergent GNAT

|| Present address: Department of Biochemistry and Molecular Genetics, University of Virginia HSC, Charlottesville, Virginia 22908, USA.

members SmAAT (ref. 7), yHAT1 (ref. 6) and AANAT (refs 8, 9), indicating a conserved role for this β -strand kink in substrate recognition by GNAT proteins.

In addition to Lys 14 of the histone substrate, Gly 13 and Pro 16 are also important for histone-binding specificity. The requirement for Lys 14 and Pro 16 stems from the extensive protein interactions described above, and the requirement for Gly 13 appears to derive from its unusually constrained packing environment within the protein and from its van der Waals interactions with Met 80 (Fig. 2b). Significantly, of the mapped acetylation sites of recombinant yGCN5 for K14 in histone H3, and for K8 and K16 in histone H4, only positions 13 and 16 (and the reactive lysine at position 14) show homology among the three sequences. Position 13 is a glycine in all three cases and position 16 is either a proline, leucine or histidine, each of which can provide hydrophobic surfaces to mimic the interactions mediated by Pro 16 of the histone H3 peptide in the complex. Thus, the histone-H3-binding determinants of tGCN5 are largely restricted to a G-K-X-P recognition sequence.

Most of the CoA-protein interactions in the tGCN5/CoA/histone H3 peptide ternary complex are made by the β 4-loop- α 3 region of the protein core domain with the pyrophosphate group and pantetheine arm of CoA, and are nearly superimposable with the interactions of acetyl-CoA in the binary tGCN5/acetyl-CoA complex (Fig. 3b), and with CoA in the hPCAF/CoA⁴ and tGCN5/CoA¹⁰ binary complexes. Additional interactions are provided by the α 1-loop of the N-terminal protein segment, and the β 4 strand and the loop- α 4 regions of the C-terminal protein segments (Fig. 4b). Most of the CoA interactions mediated by these regions in the ternary complex are rearranged relative to the binary complex, and each of

these regions is adjacent to or overlapping with protein regions that make histone H3 interactions in the ternary complex. In fact, several residues in these three regions (Leu 77 from α 1, Leu 126 from β 4, and Phe 164, Ala 165 and Phe 169 from α 4) interact with both CoA and histone H3 in the ternary complex, mostly through van der Waals interactions (Fig. 1b, c). Comparison of the tGCN5 and tGCN5/acetyl-CoA structures reveals relatively modest structural rearrangements of the tGCN5 HAT domain upon CoA binding, with most protein movements confined to the C-terminal β 5-turn- α 4 and α 5-turn- β 6 protein segments (Fig. 3a). These structural changes widen the CoA and histone substrate-binding grooves. The ternary tGCN5/CoA/histone H3 complex shows an even more pronounced rearrangement of this C-terminal protein segment and widening of the CoA and substrate-binding grooves upon histone binding. These observations indicate an intimate link between CoA and histone H3 binding, in which histone H3 binding depends on important structural contributions of CoA.

Previous enzymatic analysis of a Glu to Gln mutant at position 173 of yGCN5 (ref. 11) (corresponding to Glu 122 of tGCN5), as well as structural analysis of the related yGCN5 (ref. 3), hPCAF/CoA (ref. 4) and tGCN5/CoA (ref. 10) complexes, indicated a role for this glutamate residue as a general base for catalysis. Inspection of the active site of the ternary tGCN5/CoA/histone H3 complex (Fig. 4a) now provides a more detailed view of catalysis (Fig. 4b). Specifically, Glu 122 of tGCN5, located at the base of the substrate-binding cleft, is surrounded by several non-polar residues (Tyr 115, Phe 120, Leu 158, Tyr 160, Trp 185, Ile 189 and Leu 197) that raise the pKa of the glutamate side chain and thus facilitate its ability to extract a proton from the lysine side chain of the histone substrate. Glu 122 is

Table 1 Crystallographic data and refinement statistics

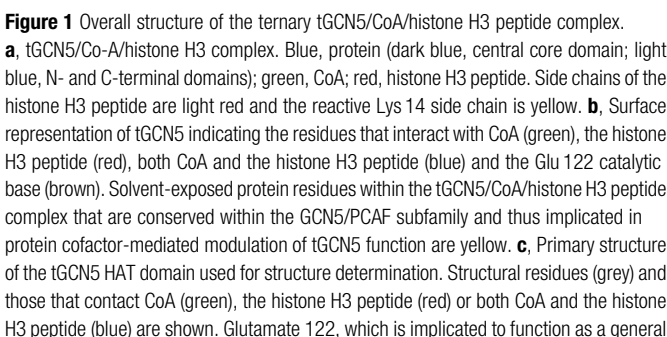
Crystal parameters							
tGCN5	$P3_2,2,1$; $a = b = 64.10$, $c = 97.8$						
tGCN5/acetyl-CoA	$P3_2,2,1$; $a = b = 64.38$, $c = 97.5$						
tGCN5/coA/histone H3	$P3_2,2,1$; $a = b = 65.09$, $c = 96.6$						
tGCN5 data collection (20.0–1.7 Å)							
Wavelength (Å)	Total reflections	Unique reflections	Completeness	R_{merge}^*	Ave I/σ		
$\lambda 1$ 0.9840	497,452	25,427	97.3	4.5%	9.4		
$\lambda 2$ 0.9795	471,314	25,223	97.8	5.4%	8.3		
$\lambda 3$ 0.9792	486,687	25,202	97.8	5.2%	8.3		
tGCN5/acetyl-CoA data collection (30.0–2.0 Å)							
Wavelength (Å)	Total reflections	Unique reflections	Completeness	R_{merge}	Ave I/σ		
1.542	158,305	16,348	99.5	5.4%	21.1		
tGCN5/CoA/histone H3 data collection (30.0–2.2 Å)							
Wavelength (Å)	Total reflections	Unique reflections	Completeness	R_{merge}	Ave I/σ		
0.917	201,683	13,391	96.9	4.7%	33.3		
Difference (to $\lambda 1$)	Phasing power†	FOM		Difference (to $\lambda 1$)	Phasing power	FOM	
		Centric	Accentric			Centric	Accentric
$\lambda 2$ disp	1.70	0.461	0.336	$\lambda 3$ disp	1.44	0.425	0.304
$\lambda 2$ anom	4.31		0.487	$\lambda 3$ anom	4.67		0.497
Combined FOM = 0.714							
Refinement statistics							
		tGCN5 ($\lambda 1$)§	tGCN5/acetyl-CoA	tGCN5/CoA/histone H3			
Resolution range		20.0–1.7 Å	30.0–2.0 Å	30.0–2.2 Å			
Number of reflections		25,403	16,307	12,590			
$R_{\text{working}}/R_{\text{free}}^\ddagger$	(%)	24.4/27.8	23.6/26.1	23.9/26.6			
RMS deviation	bonds	0.006 Å	0.006 Å	0.009 Å			
	angles	1.292°	1.335°	1.344°			
B_{average}	protein	19.98 (1,346)	35.76 (1,361)	23.98 (1,361)			
(number of atoms)	water	24.82 (96)	40.36 (72)	32.91 (122)			
	hepes	33.26 (15)	–	–			
	ACoA	–	46.46 (51)	–			
	CoA	–	–	34.61 (48)			
	Peptide	–	–	52.16 (73)			

* $R_{\text{merge}} = \sum_i |I_i - \langle I \rangle| / \sum_i I_i$.

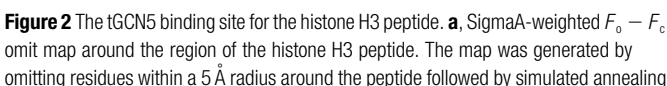
† Phasing power = $(\sum_i |F_H|^2 / \sum_i (F_{H_{\text{obs}}} - F_{H_{\text{calc}}})^2)^{1/2}$.

‡ R -factor: $R_{\text{working}} = \sum_i |F_o| - |F_c| / \sum_i |F_o|$; $R_{\text{free}} = \sum_i |F_o| - |F_c| / \sum_i |F_o|$ (where T is a test data set of 10% of the total reflections randomly chosen and set aside for refinement).

§ The MAD data set used for final refinement and phasing was from the Se-methionine-derivatized crystal.



base for catalysis, is brown, and residues that are conserved within the GCN5/PCAF subfamily of HAT enzymes (including *Arabidopsis* GCN5, *Drosophila* GCN5 and human GCN5 in addition to the alignments shown) and are solvent accessible and presumably available for additional tGCN5-mediated activities are yellow. Secondary structure elements of tGCN5 are shown above the sequence alignment. Black and grey boxes directly below the sequence alignment indicate strict and conservative conservation, respectively, of sequences within the GCN5/PCAF family of HAT enzymes. Positions of alanine mutations that decrease HAT activity are also indicated below the sequence alignment: black and grey bars indicate single¹² and triple¹³ mutations; open boxes indicate mutations that have negligible effect on HAT activity. The GNAT motifs A and D⁵ and the structurally conserved B' motif⁴ are shown.



b, Histone H3 peptide (red) interactions with tGCN5 (blue) and CoA (green).

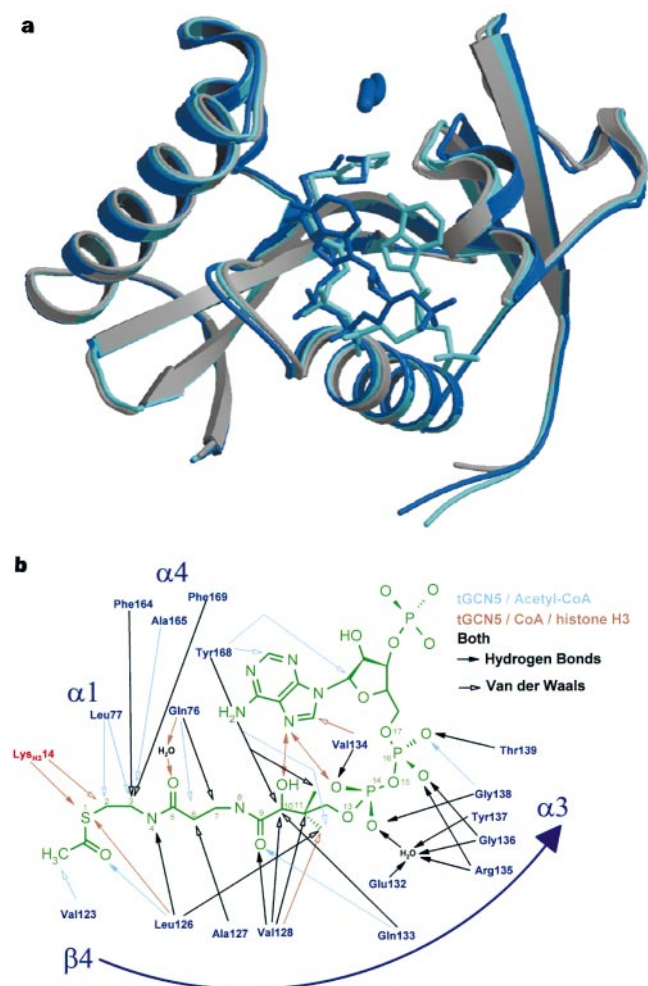


Figure 3 Comparison between tGCN5, tGCN5/acetyl-CoA and tGCN5/CoA/histone H3. **a**, Superposition of the tGCN5/CoA/histone H3 complex (blue) with tGCN5/acetyl-CoA (aqua) and tGCN5 (grey). **b**, Reorganization of tGCN5–CoA interactions upon histone H3 binding. Black arrows indicate protein residues that make conserved interactions in both the binary tGCN5/acetyl-CoA and ternary tGCN5/CoA/histone H3 complexes; brown and aqua arrows indicate residues that make interactions unique to the binary complex or to the ternary complex, respectively.

also bound to a water molecule located between its side-chain carboxylate and the reactive Lys 14 of the histone H3 peptide, and is held in place by hydrogen bonds from the backbone carbonyl of Val 123 and the backbone amide of Tyr 160 (Fig. 4a, b). This water molecule is ideally located to shuttle a proton from the reactive Lys 14 of the histone to Glu 122 of the protein; its importance is supported by its presence and relatively low B-factor in both protomers in the asymmetric unit of each of the hPCAF/CoA (ref. 4) and yGCN5 (ref. 3) structures, and in each of the tGCN5 structures presented here. The acetyl group of acetyl-CoA in the binary complex is hydrogen-bonded to the backbone amide of Leu 126, and, during acetyl group transfer to Lys 14 of the histone, this backbone amide probably polarizes the carbonyl group of the thioester before nucleophilic attack of the Lys 14 amino group and stabilizes the negative charge that develops on the oxygen in the tetrahedral transition state. Once Lys 14 of this histone substrate is acetylated, it is likely that a steric clash between the acetylsine and the backbone carbonyls of Ala 124 and possibly Val 123 causes release of the acetylated histone substrate.

The mutational sensitivity of the CoA and histone H3 peptide-binding sites correlates well with the structural observations^{12,13} (Fig. 1c). Specifically, single and triple alanine mutations that attenuate yGCN5-mediated HAT activity *in vitro* and cell growth

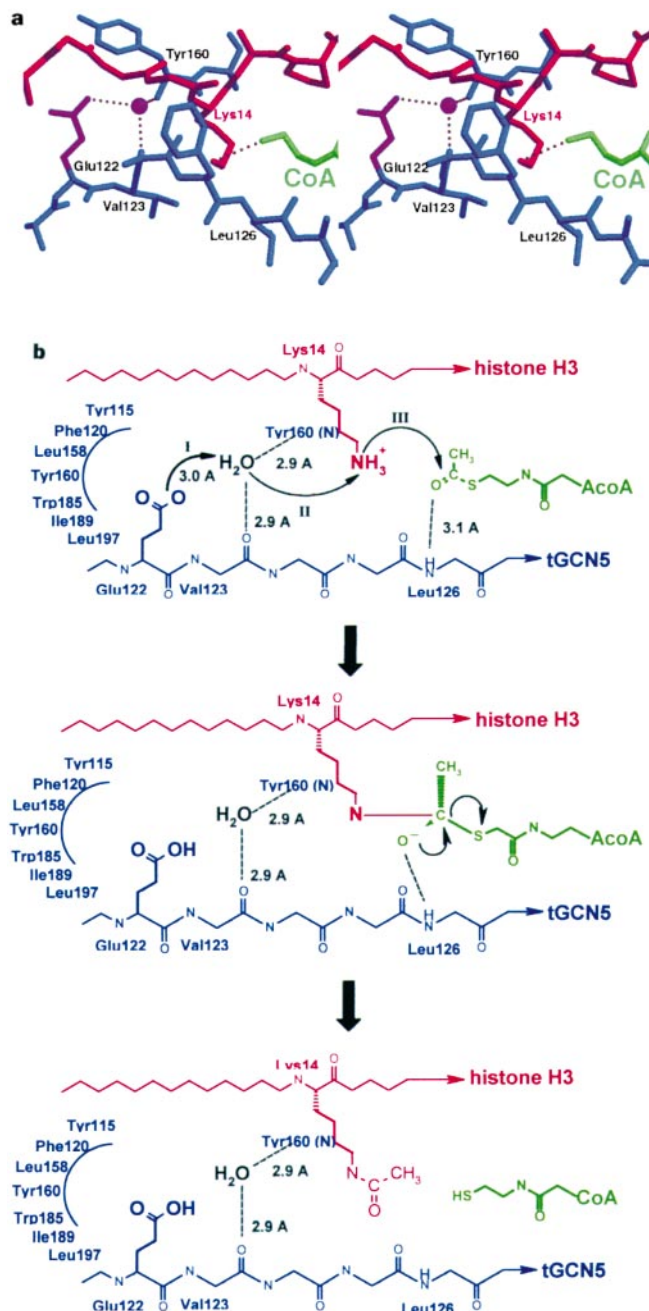


Figure 4 The tGCN5 active site. **a**, Detailed view of the active site of the ternary tGCN5/CoA/histone H3 complex. The region around histone H3 Lys 14 targeted for acetylation is shown in red. Relevant regions of CoA (green) and protein (blue) that are involved in catalysis or that make important hydrogen bonds (dotted line) or Van der Waals interactions are also shown. The putative catalytic base, Glu 122, and the water molecule proposed to play a catalytic role are purple. **b**, Proposed reaction mechanism for tGCN5-mediated catalysis of histone H3. Arrows representing proton transfer in step 1 have been consolidated for brevity. Protein residues and CoA functionalities that play a direct role in the catalytic mechanism and the Lys 14 substrate of histone H3 are indicated.

and transcriptional activation *in vivo* largely cluster around the CoA and histone H3 binding sites, and the most debilitating triple and single alanine mutations tend to affect residues that directly contact CoA or the histone H3 peptide, or both. Significantly, the mutational sensitivity of yGCN5 residues corresponding to Glu 122, Val 123 and Leu 126 of tGCN5 correlates well with their putative roles in catalysis¹¹.

In vivo, yGCN5 functions in the context of large multiprotein complexes, including SAGA and Ada, to modulate its HAT

activity^{14,15}. This observation, together with the relatively restricted histone-binding specificity of tGCN5 (for G-K-X-P), indicates that the HAT domain of GCN5 may contain protein surfaces that are contacted by protein cofactors to modulate its acetylation activity. Consistent with this hypothesis, certain residues that are conserved within the GCN5/PCAF subfamily of HAT proteins are accessible for protein cofactor interaction in the ternary tGCN5/CoA/histone H3 complex (Fig. 1b, c). These residues cluster in two distinct regions of the tGCN5 HAT domain: directly above and flanking the histone H3 binding site; and more broadly distributed over the bottom/C-terminal surface of the protein. Significantly, both regions of the protein are situated such that protein cofactor interactions may modulate tGCN5–histone interactions. Residues located above the bound histone H3 substrate may interact directly with a protein cofactor to either ‘trap’ the histone tail region in its binding site or to rearrange these residues to allow additional histone interactions, whereas protein-cofactor interactions with the C-terminal segment of tGCN5 are expected to rearrange the inherently flexible C-terminal segment for alternative histone interactions.

Comparison of the liganded forms of the tGCN5 HAT domain with the recently determined structures of the GNAT proteins yHAT1 (ref. 6), SmaAT⁷ and AANAT^{8,9}, and the N-myristoyl transferase NMT^{17,18}, shows that structurally homologous core domains provide a scaffold for their respective N- and C-terminal segments, mediate the majority of CoA contacts and harbour the residue implicated as a general base for catalysis (His 120 or His 122 in AANAT⁸, Glu 255 in yHAT1 (ref. 6), Asp 110 in SmaAT⁷ and possibly Glu 167 in NMT^{16,17}). These correlations reveal a remarkable degree of structural and functional conservation within the core domain for the large and diverse superfamily of N-acetyltransferases, possibly including other HAT proteins. The C-terminal segment of tGCN5 shows a large degree of structural variability compared with yHAT1, SmaAT, AANAT and NMT. Interestingly, the N-terminal segment of tGCN5 (β 1-loop- α 1-turn- α 2) shows structural homology with the other enzymes, even in the absence of sequence homology in this region. Significantly, the N- and C-terminal segments of tGCN5, as well as NMT and AANAT (the two other structures where substrate analogues were bound), have conserved roles in substrate binding, and in each case the N- and C-terminal domains are also involved in CoA interactions^{9,17}. Together, these correlations indicate that the GNAT proteins, and possibly other HAT proteins, use their structurally divergent N- and C-terminal segments for CoA-dependent substrate-specific binding. □

Methods

Protein expression and purification.

The HAT domain of tGCN5 (residues 48–210) was overexpressed in bacteria from a pRSETA/tGCN5 vector¹⁹ and purified to homogeneity using a combination of cation exchange, CoA affinity and gel-filtration chromatography essentially as described³. Purified protein was concentrated to ~25 mg ml⁻¹ in buffer containing 20 mM sodium citrate, pH 6.0, 150 mM NaCl and 10 mM β -mercaptoethanol, and flash-frozen and stored at -70 °C. We obtained selenomethionine-derivatized tGCN5 protein from pRSETA/tGCN5-transformed *Escherichia coli* strain B834 (DE3) (Novagen) grown in MOPS-based minimal medium supplemented with 50 mg l⁻¹ L-selenomethionine and other amino acids at the suggested concentrations; protein was isolated and stored essentially as described for the underivatized protein.

Crystallization and data collection.

We prepared crystals of the underivatized and selenomethionine-derivatized tGCN5 protein by vapour diffusion using the hanging-drop method at 4 °C. Typically, a 2.5 μ l hanging drop contained 1.5 μ l of protein (7.5 mg ml⁻¹) mixed with 1.0 μ l of reservoir containing 1.25 M Li₂SO₄ and 100 mM HEPES (pH 7.5), and equilibrated over 1 ml of reservoir solution. Binary complex crystals containing tGCN5 and acetyl-CoA were prepared by vapour diffusion using the hanging-drop method at 25 °C. The protein (7.5 mg ml⁻¹) was first mixed with acetyl-CoA (1.5 mM). A 2.5 μ l drop contained 1.5 μ l of complex and 1.0 μ l of reservoir containing 1.5 M ammonium sulfate, 50 mM HEPES (pH 7.0) and 50 mM MgSO₄, and equilibrated over 1 ml of reservoir solution. Ternary complex crystals containing tGCN5, CoA and histone H3 peptide of sequence KSTGGKAPRKQ (Baylor College of Medicine, Protein Chemistry Core Facility) were prepared by vapour

diffusion at 25 °C. The protein (7.5 mg ml⁻¹) was first mixed with CoA (1 mM) and H3 peptide (1.5 mM); 1.5 μ l of this complex was then mixed with 1.0 μ l of reservoir containing 1.6 M ammonium sulfate, 50 mM Tris (pH 7.5) and 10 mM MnCl₂, and equilibrated over 1 ml of reservoir solution. All crystals grew to a maximum size of 0.3 mm³ over about 3 days, and were transferred to a harvest buffer containing a 10% higher concentration of precipitant and salts before gradual transfer to a harvest buffer supplemented with 25% glycerol. Crystals were flash-frozen in liquid-nitrogen-condensed liquid propane before data collection.

We collected two multiwavelength anomalous dispersion (MAD) data sets (MAD_{Hg,Se} and MAD_{Se}) at 110 K from cryoprotected selenomethionine-derivatized tGCN5 crystals. The MAD_{Hg,Se} data (beamline X8C at NSLS) were collected at four wavelengths (λ 1 = 1.0093 Å (Hg edge), λ 2 = 0.9795 Å (Se edge), λ 3 = 0.9793 Å (Se peak) and λ 4 = 1.0117 Å (remote)) from a single selenomethionine-derivatized tGCN5 crystal that had been presoaked in 1.0 mM ethyl mercury phosphate for 3.0 h, and were processed using DENZO and SCALEPACK¹⁸. The MAD_{Se} data (beamline F2 at CHESS) were collected at three wavelengths (λ 1 = 0.9840 Å (Se remote), λ 2 = 0.9795 Å (Se edge) and λ 3 = 0.9792 Å (Se peak)) from a single selenomethionine-derivatized tGCN5 crystal, and were processed using MOSFLM¹⁹. A single wavelength (λ = 1.542 Å) data set was collected from a single crystal of the binary tGCN5/acetyl-CoA complex using a rotating anode and an RAXIS-IV image plate detector. A single wavelength (λ = 0.917 Å) data set was collected from a single crystal of the ternary tGCN5/CoA/histone H3 peptide complex on beamline A1 at CHESS. The data from both the binary and ternary complex crystals were processed using DENZO and SCALEPACK¹⁸. All crystals formed in identical spacegroups (P3₂,2,1) containing either one protein or protein complex per asymmetric unit cell, and formed with similar unit cell dimensions ($a = b \approx 64$ Å, $c \approx 98$ Å) (Table 1).

Structure determination and refinement.

We determined the structure of the tGCN5 protein using MAD. Briefly, the positions of all seven selenomethionine sites (excluding the N-terminal methionine) were identified from the MAD_{Se} data set using the program SOLVE (see <http://www.solve.lanl.gov>). The seven selenomethionine sites were used to phase the MAD_{Se,Hg} data set. The Hg position in this data set was determined by difference Fourier using the program PHASES²⁰. The seven selenomethionine positions and the one Hg position were then used to phase λ 1 of the MAD_{Se,Hg} data set. The resulting 1.9 Å Fourier map was improved by solvent flattening using the program PHASES²⁰. This experimental map was of excellent quality, and the program O²¹ was used to make a continuous Co trace from residues 49–208, using the Cys 200 (bound to ethyl-mercury phosphate) and the seven selenomethionine positions as guides.

Structure refinement of the tGCN5 protein was performed with the program X-PLOR²² using the λ 1 data set collected using beamline F2 at CHESS. Initial positional refinement was performed using data from 8.0 to 3.0 Å. Iterative model adjustments with O²¹ using SigmaA-weighted $2F_o - F_c$ and $F_o - F_c$ maps, and refinement with XPLOR at successively higher resolution bins of 2.7, 2.5, 2.1 and 1.7 Å, employed simulated annealing²³ and torsion angle dynamics²⁴ protocols. Towards the end of the refinement we applied a bulk solvent correction²⁵, individual atomic B-factors were refined, an ordered HEPES molecule was included and water molecules were judiciously added. The final model was checked for errors with simulated annealing omit maps²⁶. A final round of refinement at 1.7 Å resolution resulted in a protein model with excellent refinement statistics and protein geometry (Table 1).

The structures of the binary tGCN5/acetyl-CoA and ternary tGCN5/CoA/histone H3 peptide complexes were determined by molecular replacement using the program AMORE²⁷ and the refined tGCN5 structure as the search model. Structure refinement was performed with the program CNS²⁸ using the implemented maximum likelihood target function. Iterative refinement with CNS²⁸ and model building with program O²⁹ were carried out essentially as described for the tGCN5 structure. A final round of refinement of the binary tGCN5/acetyl-CoA complex at 2.0 Å resolution and the ternary tGCN5/CoA/histone H3 peptide complex at a resolution of 2.2 Å resulted in models with excellent refinement statistics and geometry (Table 1). The histone H3 peptide in the ternary tGCN5/CoA/histone H3 peptide structure encodes residues 9–19 of histone H3, although residues 9–11 encoding a Lys–Ser–Thr sequence were modelled as Ala–Ala–Gly, owing to poor electron density for the side chains of these residues.

Received 21 May; accepted 30 June 1999.

- Grant, P. A., Sterner, D. E., Duggan, L. J., Workman, J. L. & Berger, S. L. The SAGA unfolds: convergence of transcription regulators in chromatin-modifying complexes. *Trends Cell Biol.* **8**, 193–197 (1998).
- Mizzen, C. A. & Allis, C. D. Linking histone acetylation to transcriptional regulation. *Cell. Mol. Life Sci.* **54**, 6–20 (1998).
- Trievel, R. C. *et al.* Crystal structure and mechanism of histone acetylation of the yeast GCN5 transcriptional coactivator. *Proc. Natl Acad. Sci. USA* **96**, 8931–8936 (1999).
- Clements, A. *et al.* Crystal structure of the histone acetyltransferase domain of the human P/CAF transcriptional regulator bound to coenzyme-A. *EMBO J.* **18**, 3521–3532 (1999).
- Neuwald, A. F. & Landsman, D. GCN5-related histone N-acetyltransferases belong to a diverse superfamily that include the yeast SPT10 protein. *Trends Biochem. Sci.* **22**, 154–155 (1997).
- Dutnall, R. N., Tafrov, S. T., Sternglanz, R. & Ramakrishnan, V. Structure of the histone acetyltransferase Hat1: A paradigm for the GCN5-related N-acetyltransferase superfamily. *Cell* **94**, 427–438 (1998).
- Wolf, E. *et al.* Crystal structure of a GCN5-related N-acetyltransferase: *Serratia marcescens* aminoglycoside 3-N-acetyltransferase. *Cell* **94**, 439–449 (1998).
- Hickman, A., Nambodiri, M. A. A., Klein, D. C. & Dyda, F. The structural basis of ordered substrate binding by serotonin N-acetyltransferase: Enzyme complex at 1.8 Å resolution with a bisubstrate analog. *Cell* **97**, 361–369 (1999).

9. Hickman, A. B., Klein, D. C. & Dyda, F. Melatonin biosynthesis: The structure of serotonin N-acetyltransferase at 2.5 angstrom resolution suggests a catalytic mechanism. *Mol. Cell* **3**, 23–32 (1999).
10. Lin, Y., Fletcher, M., Zhou, J., Allis, C. D. & Wagner, G. Solution structure of the catalytic domain of *Tetrahymena* GCN5 histone acetyltransferase in complex with coenzyme A. *Nature* **400**, 86–89 (1999).
11. Tanner, K. G. *et al.* Catalytic mechanism and function of invariant glutamic acid-173 from the histone acetyltransferase GCN5 transcriptional coactivator. *J. Biol. Chem.* **274**, 18157–18160 (1999).
12. Kuo, M. H., Zhou, J. X., Jambeck, P., Churchill, M. E. A. & Allis, C. D. Histone acetyltransferase activity of yeast Gcn5p is required for the activation of target genes in vivo. *Genes Dev.* **12**, 627–639 (1998).
13. Wang, L., Liu, L. & Berger, S. L. Critical residues for histone acetylation by Gcn5, functioning in Ada and SAGA complexes, are also required for transcriptional function in vivo. *Genes Dev.* **12**, 640–653 (1998).
14. Grant, P. A. *et al.* Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. *Genes Dev.* **11**, 1640–1650 (1997).
15. Grant, P. A. *et al.* Expanded lysine acetylation specificity of Gcn5 in native complexes. *J. Biol. Chem.* **274**, 5895–5900 (1999).
16. Weston, S. A. *et al.* Crystal structure of the anti-fungal target N-myristoyltransferase. *Nature Struct. Biol.* **5**, 213–221 (1998).
17. Bhatnagar, R. S. *et al.* Structure of N-myristoyltransferase with bound myristoylCoA and peptide substrate analogs. *Nature Struct. Biol.* **5**, 1091–1097 (1998).
18. Otwinowski, Z. in *Proceedings of the CCP4 Study Weekend: Data collection and processing* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 56–62 (SERC Daresbury Laboratory, Warrington, UK, 1993).
19. Leslie, A. G. W. in *CCP4 and ESF-EACMB Newsletter on Protein Crystallography* (Daresbury Laboratory, Daresbury, UK, 1992).
20. Furey, W. & Swaminathan, S. in *Methods in Enzymology* (eds Carter, C. W. & Sweet, R. M.) 590–620 (Academic, Orlando, 1997).
21. Jones, T. A., Zou, J. Y. & Cowen, S. W. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
22. Brunger, A. T. *X-PLOR Manual, Version 3.8* (Yale Univ. Press, New Haven, 1996).
23. Brunger, A. T. & Krukowski, A. Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallogr. A* **46**, 585–593 (1990).
24. Rice, L. M. & Brunger, A. T. Torsion angle dynamics: Reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins* **19**, 277–290 (1994).
25. Jiang, J. S. & Brunger, A. T. Protein hydration observed in X-ray diffraction: solvation properties of penicillopepsin and neuraminidase crystal structures. *J. Mol. Biol.* **243**, 100–115 (1994).
26. Brunger, A. T., Kuriyan, J. & Karplus, M. Crystallographic R factor refinement by molecular dynamics. *Science* **235**, 458–460 (1987).
27. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
28. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
29. Jones, T. A. A graphics model building and refinement system for macromolecules. *J. Appl. Crystallogr.* **11**, 268–272 (1978).

Acknowledgements

We thank J. Berendzen, B. Sweet and their staff for access to and help on beamline X8C at NSLS; D. Thiel and his staff for access to and help on beamlines A1 and F2 at CHES; and R. Venkataramani, A. Clements, T. Stams and D. King for useful discussions. This work was supported by NIH grants to R.M., S.L.B. and C.D.A., an NIH Under-represented Minority Supplement to J.R.R., a Howard Hughes predoctoral fellowship to R.C.T. and a grant from the Fannie E. Rippel Foundation to R.M.

Correspondence and requests for material should be addressed to R.M. (e-mail: marmor@wistar.upenn.edu). Coordinates for tGCN5 (accession no. 1QST), tGCN5/acetyl-CoA (accession no. 1QSR) and tGCN5/CoA/histone H3 peptide (accession no. 1QSN) can be obtained from the Research Collaboratory for Structural Bioinformatics.